Defining the clinical and molecular spectrum of inherited eye diseases in community settings

Submitted by Siying Lin to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Medical Studies
In June 2021

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

ACKNOWLEDGEMENTS

First and foremost I would like to thank the families in this study, for their willingness to participate in research and for generously sharing their time.

I would like to thank my supervisors Prof. Andrew Crosby and Dr. Emma Baple for the opportunity to undertake a PhD and their expertise and supervision. I am grateful to Dr Barry Chioza, Dr Gaurav Harlalka and Joe Leslie for their invaluable support in the lab, and a special thanks to Dr Mark Gilchrist for his pastoral support. I have had the opportunity to collaborate with many wonderful scientists and clinicians in the course of my PhD, and I would like to thank them all for their contributions.

I am fortunate to have worked with colleagues who have become close friends, including Ilaria D'Atri, Olivia Rickman, Hannah Jones and Joe Leslie; thank you for keeping me sane. Special thanks to my good friends Annika Quinn, Roselin Charles and Natalee James for doing the same.

To my long-suffering parents and sister Sim, thank you for your unwavering support and encouragement, I could not have completed this journey without you. Thank you to Charlie and Laurence for cheering me on at the finishing line.

The work in this thesis was funded by the University of Exeter Vice Chancellor Scholarship.

ABSTRACT

Inherited eye diseases are an important contributor to the burden of childhood blindness globally. These conditions are often associated with significant phenotypic and genetic heterogeneity and are extremely difficult to study in a general population setting. This thesis details the study of inherited eye diseases in genetically isolated populations including the North American Amish and rural Pakistani and Palestinian communities. Here, an enrichment of disease-causing founder mutations arising from common ancestry, characteristic marriage patterns and geographical isolation, combined with the often large family sizes typical of families in these regions, enables powerful genomic studies to identify pathogenic sequence variants. As well as providing an important opportunity to learn about the genetic causes of inherited eye diseases, these studies also provide desperately required healthcare benefits for the families and populations involved.

Chapter 3 describes studies of oculocutaneous albinism (OCA) in 40 Amish and Pakistani families. Results from comprehensive clinical, genomic and functional studies, initiated by a search for the cause of OCA in a number of Amish families, provide strong evidence for the pathogenicity of two common *TYR* gene variants [p.(Ser192Tyr) and p.(Arg402Gln)] when inherited *in cis*. These variants were previously considered gene polymorphisms although this has been heavily debated in many studies, and these variants are currently variably reported and even potentially excluded by clinical testing laboratories. The findings reported in this thesis have important diagnostic implications by helping clarify the contribution of these variants to the OCA phenotype, and by promoting the reporting the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) in *cis* haplotype as a pathogenic allele. This will likely increase the molecular diagnoses in albinism patients with missing heritability by 25-50%. This chapter also entails a comprehensive investigation involving genetic studies alongside an exhaustive literature review of all published OCA genetic causes in Pakistani families, including cross-referencing with established genomic databases to evidence the likely causality of each gene variant.

Chapter 4 entails clinical and genomic findings in four families with phenotypic features highly suggestive of a ciliopathy disorder. Findings include identification of novel *SCAPER* and *BBS5* variants, enabling a more precise definition of the SCAPER clinical phenotype. This work also consolidates an *INPP5E* c.1879C>T; p.(Gln627*) variant, a likely pathogenic founder alteration present in Northern Pakistan, as a cause of MORM syndrome.

Chapter 5 documents studies of families with rare and ultra-rare inherited ocular diseases in Pakistani and Palestinian communities. This includes consolidating *SDHD* dysfunction as a cause of mitochondrial disease through investigations of an extended Palestinian family, facilitating a clearer delineation of the variable ocular (and non-ocular) phenotypical features. Alongside this, the identification of novel and known variants in *ALDH1A3*, *FYCO1*, *TDRD7*, *CYP1B1*, *ATOH7*, *LRP5*, *FRMD7* and *HPS1* in Pakistani communities contributes to an improved knowledge of the genetic spectrum and frequencies of various forms of inherited eye diseases regionally.

The comprehensive OCA and BBS datasets provide notably improved knowledge, as well as a centralised repository, of the genetic spectrum and regional frequencies of the molecular causes of these conditions in the Amish and in Pakistan. Ultimately, these findings will greatly facilitate the establishment of robust cost-effective accurate diagnostic genetic testing, clinical management and counselling efforts. Together the work of this thesis describes data of scientific importance, and highlights the immense value of translational community research studies to deliver clinical benefits locally and globally in the field of inherited eye diseases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

<	Less than
>	Greater than
≤	Less than or equal to
≥	Greater than or equal to
°C	Degrees celsius
~	Approximately
AAV	Adeno-associated viral
ADHD	Attention deficit hyperactivity disorder
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BAM	Binary Alignment/Map
BBS	Bardet-Biedl syndrome
bHLH	Basic helix-loop-helix
BLAT	BLAST-like alignment tool
BN-PAGE	Blue native polyacrylamide gel electrophoresis
bp	Basepair
BWA	Burrows-Wheeler Aligner
$CaCl_2$	Calcium chloride
cDNA	complementary DNA
CHARGE	Coloboma, heart defects, atresia choanae, growth retardation, genital abnormalities, ear abnormalities
CLIA	Clinical Laboratory Improvement Amendments
CNS	Central nervous system
CNV	Copy number variant
CO_2	Carbon dioxide
CRISPR	Clustered Regularly Interspersed Palindromic Repeat
dATP	Deoxyadenosine triphosphate
dbSNP	Single Nucleotide Polymorphism Database
dCTP	Deoxycytidine triphosphate
ddH2O	Double-distilled water
dGTP	Deoxyguanosine triphosphate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DP	Read depth
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
Exo I	Exonuclease I
FEVR	Familial exudative vitreoretinopathy
FHONDA	Foveal hypoplasia, optic nerve decussation defects, and anterior segment dysgenesis
FMS	Illumina fragmentation solution
FOX	Forkhead box

g	Gram
GATK	Genome Analysis Toolkit
GC	Guanine-cytosine
GDPR	General Data Protection Regulation
g-force	Relative centrifugal force
gnomAD	Genome Aggregation Database
GRCh37/38	Genome Reference Consortium human genome build 37/38
HEK293F	Human Embryonic Kidney 293 Freestyle
HGMD	Human Gene Mutation Database
HPS	Hermansky-Pudlak syndrome
HTA	Human Tissue Authority
IAPB	International Agency for the Prevention of Blindness
IC3D	International Committee for Classification of Corneal Dystrophies
IGV	Integrative Genome Viewer
Indel	Insertion and deletion
iPSC	Induced pluripotent stem cell
JBTS	Joubert syndrome
kb	Kilobase
kDa	Kilodalton
KPK	Khyber Pakhtunkhwa
LAB	Lithium acetate borate
LB	Lysogeny broth
LogMAR	Logarithm of the Minimum Angle of Resolution
LRVC	Low vision resource centre
MAF	Minor allele frequency
Mb	Megabase
$MgCl_2$	Magnesium chloride
MIM	Mendelian Inheritance in Man
min	Minute
ml	Millilitre
mm	Millimetre
mM	Millimolar
MORM	Mental retardation, truncal obesity, retinal dystrophy and micropenis
MQ	Mapping quality
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
N	Normality
NAD	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
ng	Nanogram
NGS	Next generation sequencing
nm	Nanometer
NNSPLICE	Splice site prediction by neural network
NPHP	Nephronophthisis
OCA	Oculocutaneous albinism
OCT	Ocular coherence tomography
OMIM	Online Mendelian Inheritance in Man

ONT	Oxford Nanopore Technologies
OXPHOS	Oxidative phosphorylation
PacBio	Pacific Biosciences
PB1	Illumina Prepare BeadChip Buffer 1 solution
PCD	Primary ciliary dyskinesia
PCR	Polymerase chain reaction
PH-like	Pleckstrin Homology-like
PM1	Illumina precipitation solution
PMSF	Phenylmethylsulfonyl fluoride
PolyPhen-2	Polymorphism Phenotyping v2
PROVEAN	Protein Variation Effect Analyser
PTC	Premature termination codon
RA1	Illumina resuspension, hybridisation and wash solution
RGC	Retinal ganglion cell
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RPE	Retinal pigment epithelium
rpm	Revolutions per minute
rSAP	Shrimp alkaline phosphatase
SAM	Sequence Alignment/Map
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Seconds
SIFT	Sorting Intolerant From Tolerant
SMRT	Single-molecule real-time
SNP	Single Nucleotide Polymorphism
S.O.C	Super Optimal broth with Catabolite repression
SSF	SpliceSiteFinder-like
Ta	Annealing temperature
TRID	Translational read-through inducing drug
TUDCA	Tauroursodeoxycholic acid
U	Units
UCSC	University of California, Santa Cruz
UK	United Kingdom
USA	United States of America
v	Version
V	Volt
VCF	Variant Call Format
WES	Whole exome sequencing
WoH	Windows of Hope
WHO	World Health Organisation
µl	Microlitre
µM	Micromolar

1 INTRODUCTION

1.1 Anatomy of the eye

The eye is a highly specialised sensory organ which enables visual function through photoreception and phototransduction. Light from the environment is absorbed by specialised rod and cone photoreceptors in the retina, which converts light stimulus into nerve action potentials. These electrical impulses are transmitted via bipolar cells to ganglion cells, whose long axonal fibres converge to form the optic nerve, optic chiasm and optic tract. This neural information is subsequently relayed to the visual cortex in the brain, where it is processed and consciously appreciated as vision. Other structures in the eye serve to support this basic physiological process, either by helping to focus and transmit light to the retina, such as the cornea, lens, iris and ciliary body, or to nourish and support the tissues of the eye, such as the choroid, aqueous outflow system, and the lacrimal apparatus.

In humans, the eye is roughly spherical in shape with an approximate diameter of 23 mm. It is made up of three basic layers of tissue; the outer fibrous corneoscleral layer, the middle uveal layer, and the inner retinal layer (Figure 1.1). The cornea and sclera together form a tough fibrous envelope that protects and supports the ocular tissues. The sclera is opaque and provides structural support for attachment of the extraocular muscles, whilst anteriorly, the transparent cornea serves as the principle refracting medium of the eye and roughly focuses an image onto the retina. The border of the cornea and the sclera is known as the limbus. The middle layer, the uvea or uveal tract, is composed of the choroid, the ciliary body and the iris. The choroid is a thin, heavily pigmented, vascular connective tissue that lies between the sclera and the retina. It nourishes and supports the outer retinal layer, and aids vision by absorption of scattered light. Anteriorly, the choroid merges with the ciliary body, which is a circumferential thickening of the uvea lying beneath the limbus. It consists of the ciliary epithelium and the ciliary muscle, which are important in aqueous humour production and accommodation respectively.

The eye is divided into two segments by the ciliary body, zonules and lens. The smaller anterior segment is filled with aqueous humour, which is secreted by the ciliary epithelium, and helps maintain the intraocular pressure within the eye as well as providing nutrition to the avascular cornea and lens. The larger posterior segment contains the transparent viscoelastic vitreous gel that helps maintain the shape of the eye. The heavily pigmented iris lies anterior to the ciliary body and lens, and contains a central aperture, the pupil. The iris incompletely divides the anterior segment into the larger anterior chamber and the smaller posterior chamber, which communicate through the pupil. The contractile iris acts as an adjustable diaphragm by altering the pupil size, thereby regulating the amount of light reaching the retina. The crystalline lens is a transparent, biconvex structure that is attached to the ciliary body via radially arranged suspensory ligaments or zonules. Changes in the tone of the ciliary muscle alters the tension of the zonules, resulting in changes to the shape of the naturally elastic lens. This then changes the refractive power of the lens, facilitating accommodation by allowing the eye to focus from distant to near images.

The innermost layer of the eye is the photosensitive retina, which consists of two primary layers, an inner neurosensory retina, and an outer retinal pigment epithelium (RPE). The retina terminates anteriorly behind the ciliary body in a scalloped line known as the ora serrata. The macula is an oval-shaped pigmented area near the center of the retina, and this area is responsible for central high-acuity and colour vision. The visual axis of the eye passes through the fovea centralis, a depression in the center of the macula where cone photoreceptors are concentrated at the maximum density to the exclusion of rod photoreceptors. The optic nerve is formed by convergence of retinal ganglion cell (RGC) axons at the optic disc and exits the eye through a sieve-like opening in the sclera known as the lamina cribrosa.

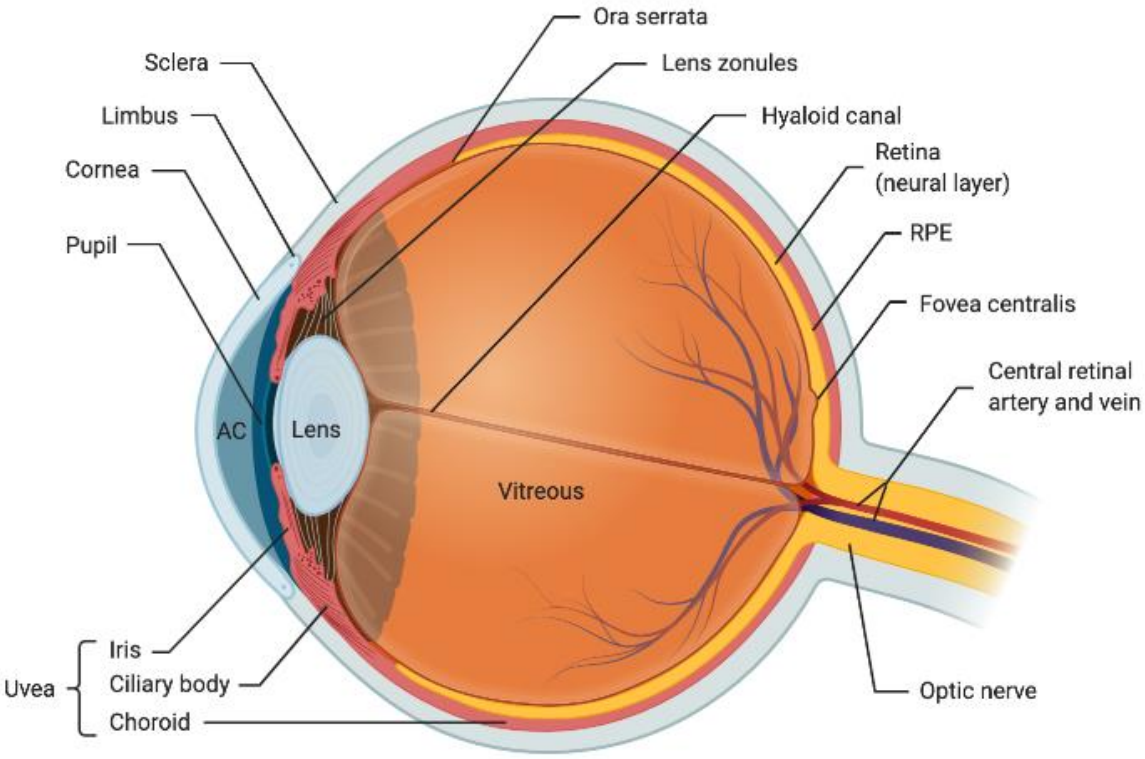


Figure 1.1 Anatomy of the human eye

Abbreviations: AC, anterior chamber; RPE, retinal pigment epithelium. Adapted from Koeppen et al *(1), created with BioRender.*

1.2 Ocular development and malformations

1.2.1 General ocular development

The adult eye is a complex organ with multiple tissue components and cell types, each with diverse yet specific functions. The development of the eye can be empirically subdivided into three phases. The first phase involves induction through localised signalling and regional specification to form the major morphological structures of the eye. The second phase involves cellular differentiation and maturation of these structures to form the functional eye. The third phase then involves the formation of neuronal connections between the retina and the visual processing centre in the brain (2). Normal ocular development is a coordinated multi-step process that is regulated through a complex interaction of signalling molecules and transcription factors encoded for by a cascade of genes that are activated in specific cell types in a specific order. Vision and ocular development genes are highly conserved, suggesting similar

genetic developmental programs across species. One such gene is *PAX6*, which encodes a transcription factor that is a key regulator of several ocular developmental processes (3). The amino acid sequence of the human Pax6 protein is identical to the mouse protein (4) and shares 97% sequence identity to the zebrafish protein (5). Mutation of the *PAX6* gene in humans causes aniridia, characterised by partial or complete absence of iris tissue (6), whilst mutations in the mouse *Pax6* gene, known as small-eye (Sey), cause iris hypoplasia and other ocular defects similar to human aniridia (7). Developmental eye anomalies such as anophthalmia and microphthalmia can also be caused by pathogenic mutations in a number of other ocular developmental genes including *SOX* and *OTX2* (8).

The developing embryo consists of three germinal layers: the ectoderm, the mesoderm and the endoderm. The ectoderm then differentiates in the outer surface ectoderm and the inner neuroectoderm. Developing vertebrate embryos also have a multipotent stem cell population, the neural crest cells, which migrate and can differentiate into diverse cell types. Ocular and orbital tissues develop from the neuroectoderm, with contributions from the surface ectoderm, mesoderm, and significantly from the neural crest cells. Neural crest cells also give rise to many facial and skull structures, and therefore ocular and craniofacial anomalies are often seen in syndromes such as Goldenhar syndrome and CHARGE syndrome that arise from neural crest maldevelopment (9).

In humans, eye development begins during gastrulation, when the eye field, or the eye-forming region, is induced within the anterior part of the neural plate (derived from neuroectoderm). The eye field later splits along the midline during neurulation, and the optic grooves develop, which later form the right and left optic vesicles (Figure 1.2). A key inductive signal in this process is the sonic hedgehog protein encoded for by the *SHH* gene, and failure of this process results in a midline defect known as holoprosencephaly, associated with wide ocular phenotypic variability ranging from isolated coloboma, microphthalmia, cyclopia, or complete anophthalmia (10).

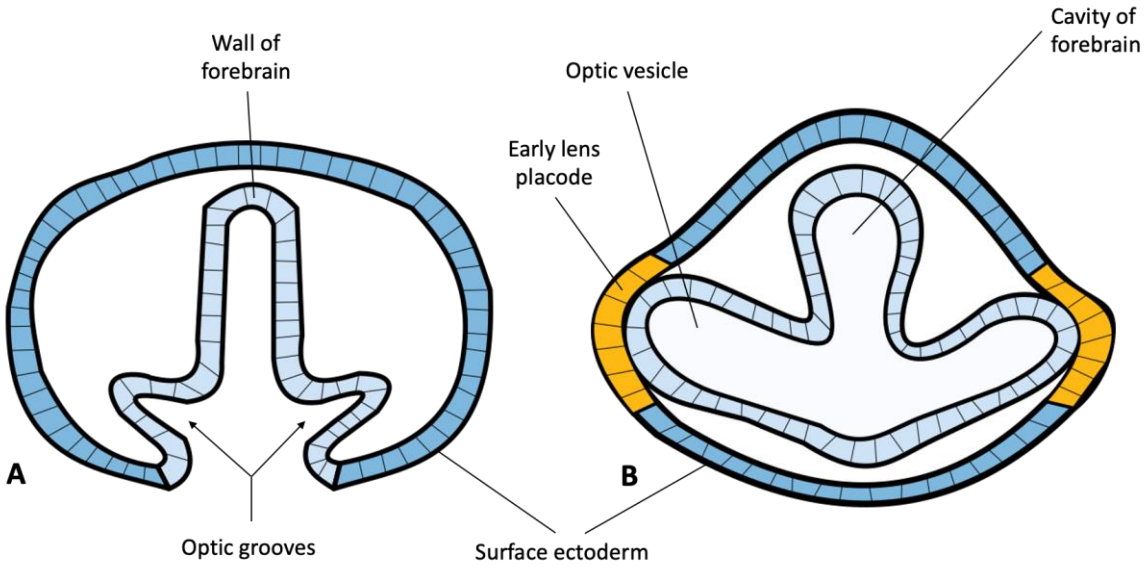


Figure 1.2 Development of the optic vesicles

(A) Transverse section through developing forebrain at day 22 showing development of the optic grooves.
(B) Transverse section through developing forebrain at week 4; contact between the neuroepithelium of the optic vesicles and the surface ectoderm induces the formation of the lens placode.
Adapted from Lamb (11).

The optic vesicles interact with and induce the overlying surface ectoderm to thicken into the lens placode. The distal part of the optic vesicle then invaginates, forming a bilayered optic cup. The inner layer of the optic cup will form the neural retina, whilst the outer layer differentiates into the RPE (Figure 1.3). Invagination of the optic cup occurs asymmetrically, with a groove, the optic fissure, developing from the ventral side of the optic cup into the optic stalk. This fissure facilitates the entry of hyaloid vessels that supply nutrients to the inner layer of the optic cup and the lens vesicle during ocular development. Closure of this fissure occurs around week 5 of gestation, and starts in the centre of the fissure, before “zipping up” both anteriorly and posteriorly. The optic fissure closes by week 6, establishing the basic structure of the eye; incomplete closure results in a coloboma, an ocular malformation that can affect one or several structures in the eye including the iris, choroid, retina or optic disc. Colobomas are most commonly located in the inferonasal quadrant of the eye, which correlates with the location of optic fissure closure (12).

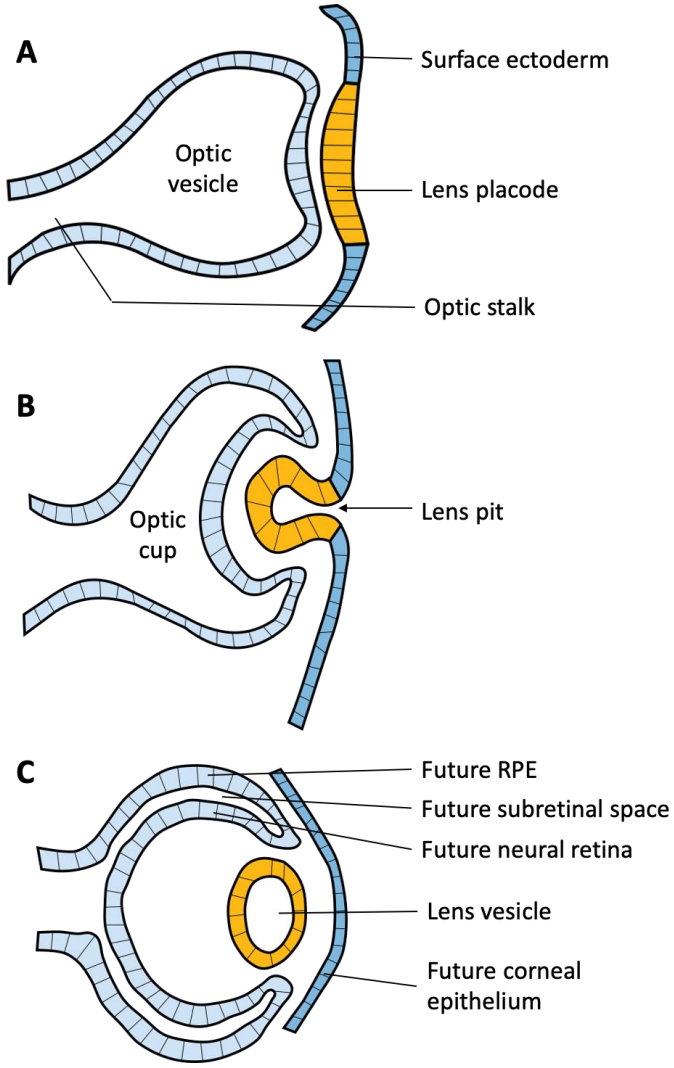


Figure 1.3 Development of the lens vesicle and optic cup

(A) Development of the optic vesicle and localised thickening of the surface ectoderm to form of the lens placode.
(B) Invagination of the lens vesicle into the optic cup.
(C) Lens pit deepens and detaches from the ectoderm to form a lens vesicle; invagination of the optic vesicle forms a bilayered optic cup which will eventually develop into the neural retina and retinal pigment epithelium (RPE) layers of the retina. Adapted from Erskine et al (13)

In parallel to the development of the optic cup, the lens placode invaginates into the optic cup, forming a shallow depression known as the lens pit. The lens pit deepens, closes anteriorly and detaches from the surface ectoderm, forming the lens vesicle (Figure 1.3). Delayed or incomplete separation of the lens vesicle from the surface ectoderm can result in an ocular malformation known as Peters anomaly, associated with central or paracentral corneal opacities (leukoma), absence of the adjacent posterior corneal stroma, corneal endothelium and descemet's membrane, and a

variable degree of iris and lenticular adhesions to the posterior cornea. Most cases of Peters anomaly are sporadic without an identified genetic cause, although some cases have been associated with mutations in *MAF*, *FOXC1*, *FOXE3*, *PITX2* and *PAX6* (14).

As the optic vesicles extend outward, their proximal attachment to the developing forebrain becomes the optic stalk. Growth of RGC axons into the optic stalk in week 7 leads to the formation of the optic nerve, connecting the eye with the visual centres of the brain (12).

1.2.2 Development of the lens

The lens vesicle is a symmetrical unicellular epithelial structure that is almost spherical in shape, and surrounded by a basal lamina layer that will eventually form the lens capsule. Growth factors from the developing neural retina induce the posterior cells of the lens vesicle to elongate, forming the primary lens fibres. The base of each elongating lens cell remains anchored to the basal lamina posteriorly, whilst their apices grow anteriorly, thereby filling the central cavity of the lens vesicle. The anterior lens epithelial cells migrate from the central to the equatorial regions of the lens vesicle, where they undergo mitosis and differentiate to form secondary lens fibres. These surround the primary lens fibres in concentric layers in a process that continues into adulthood, resulting in an ellipsoid-shaped lens. Differentiated lens fibre cells no longer contain nuclei or mitochondria, and accumulate specialised proteins known as crystallins, leading to lens transparency.

Lens development is a complex and delicate process and susceptible to disruption, resulting in congenital lens opacities or cataracts. Over half of all congenital cataracts are idiopathic (15); identified causes of congenital cataracts include genetic aetiologies, intrauterine infections and teratogen exposure. Congenital cataracts due to genetic aetiologies can be isolated (with autosomal dominant, autosomal recessive and X-linked inheritance patterns described), associated with other ocular anomalies such as aniridia, anterior segment dysgenesis, persistent foetal vasculature, and posterior lenticonus, or associated with a number of chromosomal abnormalities, genetic syndromes or metabolic diseases (16). Inherited congenital cataracts have been associated with pathogenic variants in genes encoding lens proteins including

crystallins (*CRYAA*, *CRYAB*, *CRYBA1*, *CRYBA2*, *CRYBA4*, *CRYBB1*, *CRYBB2*, *CRYBB3*, *CRYGB*, *CRYGC*, *CRYGD* and *CRYGS*), cytoskeletal structural proteins (*BFSP1* and *BFSP2*) and membrane proteins (*GJA3*, *GJA8*, *MIP*, *LIM2*, *EPHA2*, and *DNMBP*), as well as the transcription factors *FOXE3*, *PITX3*, *HSF4* and *MAF* (17).

1.2.3 Development of the cornea, iris and ciliary body

After separation of the lens vesicle, the surface ectoderm reforms. The underlying lens vesicle then induces the overlying surface ectoderm to differentiate into the corneal epithelium. Neural crest cells migrate to this region in three waves. The first wave of cells migrate into the space between the surface ectoderm and the anterior surface of the lens and give rise to the corneal endothelium and trabecular meshwork. A second wave of cells then migrates between the corneal epithelium and endothelium to form the corneal stroma. Later, a third wave of cells migrate to the space between the corneal endothelium and the anterior edge of the optic cup, contributing to the ciliary body and iris stroma (18). The iris and ciliary body also develop from the margins of the optic cup under inductive influences from the developing lens, with the optic cup margin ultimately forming the pupil margin.

FOXC1 and *PITX2* are genes that encode transcription factors with critical roles in regulating the process of neural crest cell migration in the developing eye. Both *Foxc1* and *Pitx2* are co-expressed in periocular mesenchyme (derived from neural crest cells and mesoderm mesenchyme) in the developing mouse eye, and physically interact through crucial functional domains (19). In humans, mutations in *FOXC1* or *PITX2* cause Axenfeld-Rieger syndrome, an autosomal dominant disorder characterised by anterior segment dysgenesis and systemic abnormalities. Ocular features of Axenfeld-Rieger syndrome mainly affect the iris, cornea and anterior chamber angle, structures that are derived from the periocular mesenchyme, and include iris hypoplasia, corectopia, polycoria, posterior embryotoxon, and iris strands connecting the iridocorneal angle to the trabecular meshwork (20).

1.2.4 Development of the retina and RPE

All retinal tissues develop from the bilayered optic cup, with the neural retina developing from the inner layer of the optic cup, whilst the outer layer forms the RPE. As a result of the invagination process, the apical surface of the neural retina lies adjacent to the apical surface of the RPE, with a potential space between the two layers derived from the cavity of the optic cup; this intra-retinal space later disappears as the two layers fuse. There are no real anatomic junctions formed between the cells in apposition in the two layers (photoreceptors of the neural retinal and RPE cells) (21). Therefore, whilst this interaction between the layers is crucial for the functional maintenance of the photoreceptors, the attachment between layers remains weak, and can be overcome by a number of mechanisms, resulting in a retinal detachment and re-establishment of the potential intraretinal space between these two layers.

Differentiation of the retinal layers and cell types proceeds in a highly regulated manner, starting at the posterior pole and proceeding peripherally in a concentric manner; a gradient of retinal differentiation can therefore be seen within an individual eye. Retinal vasculature follows the same concentric pattern of development, arising from the hyaloid circulation at the optic disc and spreading peripherally, first reaching the nasal periphery at 8 months gestation, whilst the temporal periphery is not completely vascularised till shortly after birth at full term. As such, temporal retinal vascularisation may not be fully developed in premature infants, explaining the greater susceptibility of the temporal retina to retinopathy of prematurity, a retinal vasoproliferative disorder affecting premature neonates (22). Congenital anomalies in retinal vascular development are also associated with inherited ocular diseases such as familial exudative vitreoretinopathy (FEVR), characterised by incomplete vascularisation of the peripheral retina and poor vascular differentiation, which can lead to complications such as the development of fibrovascular membranes resulting in vitreoretinal traction, retinal folds and retinal detachment, and potentially profound vision loss. Mutations in at least 10 genes have been associated with the FEVR phenotype including *NDP*, *FZD4*, *LRP5*, *TSPAN12*, *ZNF408*, *CTNNB1*, *KIF11*, *RCBTB1*, *JAG1* and *ATOH7* (23, 24); of these, *NDP*, *FZD4*, *LRP5* and *TSPAN12* have been shown to play a role in the Norrin/Frizzled4 signalling pathway, suggesting a crucial role for this pathway in retinal vascular development (23).

The primitive neural retina consists of two zones, an inner non-nucleated ‘marginal zone’ and an outer nucleated ‘primitive zone’. At around week 7 of gestation, newly formed cells from the mitotically active primitive zone migrate into the marginal zone to form the inner neuroblastic layer; the outer nucleated zone is now known as the outer neuroblastic layer. The inner neuroblastic layer eventually gives rise to the Müller cells, ganglion cells and amacrine cells, whilst the outer neuroblastic layer gives rise to bipolar cells, horizontal cells and primitive photoreceptor cells. The inner and outer neuroblastic layers are surrounded by the internal and external limiting membranes respectively, which form when the cells within these layers cease to proliferate and start differentiating. Lamination of the neural retina occurs at approximately 8-12 weeks of gestation, with ganglion cells being the first cells to differentiate. Later, Müller cells, amacrine cells, bipolar cells, horizontal cells, and photoreceptors are formed, and their cell bodies and synaptic connections contribute to the nuclear and plexiform layers seen in the familiar 10-layered retinal structure (Figure 1.4).

Each retinal cell type is not determined by an invariant cell lineage, instead, each retinal progenitor cell appears to have the potential to form any of the mature retinal neurons or glial cell types (25). This neural retinal development is driven by overlapping cascades of genetic programs that help determine cell lineages and commitment to specific cell fates, regulated by the temporal and spatial expression of transcription factors in the retina encoded by three major groups of genes; the basic helix-loop-helix (bHLH), the forkhead box (FOX) and the homeobox genes (26). Variants in retina-specific transcription factors such as *NR2E3* are associated with primary retinal disorders (27), whilst variants in more widely expressed transcription factors can result in systemic disorders with ocular manifestations; for example, variants in *NEUROD1*, a bHLH transcription factor that functions in brain and pancreas development, are associated with neonatal diabetes with retinal degeneration (28).

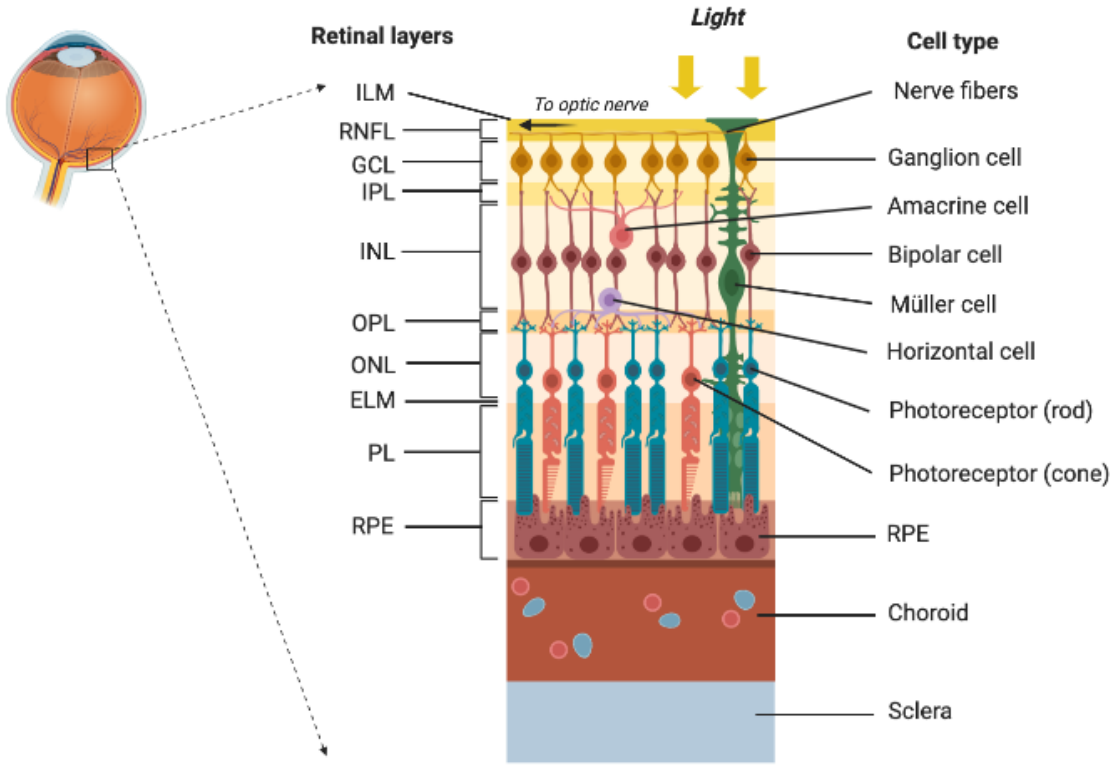


Figure 1.4 Cell types and histologic layers in the adult human retina

The internal limiting membrane (ILM) is the innermost boundary of the retina and is formed by the footplates of the Müller cells. The retinal nerve fibre layer (RNFL) is formed by the axons of the ganglion cells, whilst the ganglion cell layer (GCL) contains ganglion cell nuclei. The inner plexiform layer (IPL) is made up of the neuronal processes and synapses of bipolar, ganglion and amacrine cells, whilst the outer plexiform layer (OPL) is composed of the interconnections between the photoreceptor synaptic bodies and the horizontal and bipolar cells. The inner nuclear layer (INL) contains bipolar, amacrine, horizontal and Müller cell nuclei, whilst the outer nuclear layer (ONL) contains the cell bodies of the rod and cone photoreceptors. The external limiting membrane (ELM) is the outermost layer of the neurosensory retina; it is formed by the attachment sites of adjacent photoreceptors and Müller cells, and separates the photoreceptor inner segments and cell nuclei. The photoreceptor layer (PL) contains the inner and outer segments of the rod and cone photoreceptors. The retinal pigment epithelium (RPE) consists of a single layer of pigmented epithelial cells. Adapted from Koeppen et al (1), created with BioRender

Photoreceptors arise from the outermost cells of the primitive neural retina. Cone photoreceptors develop first, followed by rod photoreceptors. The outer segments of photoreceptors, which are actually modified ciliated processes, consist of stacks of membranous discs containing opsin, a visual pigment that is key to initiation of the phototransduction cascade. These outer segments are in close contact with the RPE. Mutations in *NRL* and *CRX,* which encode transcription factors that regulate photoreceptor differentiation during retinal development, have been shown to retinal dystrophy phenotypes in humans (29, 30).

Development of the macula differs from the rest of the retina. The macula is first evident as a localised increase in ganglion cell density temporal to the optic disc. This is later followed by displacement of the ganglion cells and the formation of an area of localised thinning, the foveal depression. The foveal structure is still immature at birth, and it can take up to 15 - 45 months post-partum for the fovea to reach full histological maturity, where the inner nuclear and ganglion cell layers have receded to the margins of the fovea, leaving only densely packed cone photoreceptors in the foveal region (31).

At birth, the ganglion cell layer at the fovea is still a single cell thick; in fact, full macula development is not complete till 4 months postpartum. In foveal hypoplasia, the fovea appears to be poorly developed or absent. This feature in isolation is rare, and is usually described in association with other ocular disorders such as albinism, aniridia, microphthalmia, retinopathy of prematurity, achromatopsia, optic nerve hypoplasia, FEVR and Stickler syndrome (32, 33). Studies of inherited disease phenotypes associated with foveal hypoplasia have identified a number of genes and molecular pathways critical for normal foveal development, including the *PAX6* transcription factor and members of the Wnt signaling pathway (*FZD4* and *NDP*) (34).

The RPE develops from the outer layer of the optic cup, and is initially comprised of a single layer of ciliated pseudostratified columnar epithelial cells; the cilia later disappear as melanogenesis commences. The basement membrane of the RPE forms one of the five layers of the Bruch's membrane, and tight junctions between the retinal pigment epithelial cells form the outer blood-retinal barrier.

1.2.5 Development of vitreous and hyaloid system

The vitreous develops in the space between the inner layer of the optic cup and the lens vesicle in three distinct stages. At 5 weeks gestation, the lentoretinal space is occupied by the primary vitreous. This is derived from mesenchymal cells that have entered the optic cup via the optic fissure, as well as from lens surface ectoderm cells and neuroectoderm cells from the inner layer of the optic cup (35). The hyaloid artery, which is a branch of the primitive ophthalmic artery, enters the optic cup via the optic fissure and courses through the primary vitreous to reach the posterior lens surface, where it branches to form a network of capillaries, the tunica vasculosa lentis.

The secondary vitreous is derived entirely from the neuroectoderm, and is composed of fine, organised fibrillary material (35). It is initially deposited at the posterior pole behind the primary vitreous, later extending to envelop the entire primary vitreous. Tertiary vitreous refers to a phase of development around the end of the third month, where distinct condensations of the secondary vitreous become evident at the lens equator, associated with the formation of lens zonules, and become firmly attached to the inner limiting membrane of the retina in the developing pars plana region, forming the vitreous base. The tertiary vitreous and zonular fibres are actually thought to be produced by the neuroectoderm of the developing ciliary body (35).

During the 4th month of gestation, regression of the primary vitreous, the hyaloid vessels, and the tunica vasculosa lentis occurs. The course of the regressing hyaloid artery is evident in the adult vitreous as Cloquet's canal, a narrow, fluid-filled central channel that extends from the optic disc posteriorly to the region of the degenerating tunica vasculosa lentis anteriorly. Failure of the hyaloid system to regress can result in developmental abnormalities of the vitreous such as a Mittendorf dot, Bergmeister's papilla, persistent hyaloid artery, and persistent fetal vasculature (previously known as persistent hyperplastic primary vitreous) (35). The adult vitreous is a specialised extracellular matrix composed of hyaluronan, the major macromolecule, interwoven with collagen fibrils (composed of collagen types II, IX, and V/XI) (36). As such, inherited diseases resulting from mutations in genes affecting collagen or hyaluronan function, such Marfan syndrome (*FBN1*), Ehlers-Danlos syndrome (*COL1A1*, *COL1A2*, *COL5A1* and *COL5A2*), Stickler syndrome (*COL2A1*, *COL9A1*, *COL11A1*),

Knobloch syndrome (*COL18A1*), Wagner syndrome (*VCAN*) and HYAL2 syndrome (*HYAL2*), also often display ocular features associated with vitreoretinal degeneration, including vitreous syneresis, vitreous traction, high myopia and retinal detachment (37, 38).

1.3 Genetic studies in communities

The field of community genetics or genomics can be defined as "the art (application) and science (research) of the responsible and realistic application of health and disease-related genetics and genomics knowledge and technologies in human populations and communities to the benefit of individuals therein" (39). Communities can be defined by their geographical location, ethnic origins, or by their cultural, religious or socio-economic characteristics. Community genetics aims to bring about clinical benefits to individuals and families affected by genetic diseases, as well as to locate individuals within the wider community who may also be at risk of the same inherited conditions. Communities also benefit through liaison with support organisations and health authorities to ensure delivery of up-to-date genetic counselling, screening and clinical care to affected individuals.

Our research team has established collaborations with scientists and healthcare professionals in several communities globally. There is a strong emphasis on translating the scientific insights arising from our research studies into improved healthcare outcomes for communities as well as affected individuals globally through new therapies and diagnostic strategies as well as refinement of genomic healthcare policy. This thesis focuses on inherited eye diseases in affected families from Amish, Pakistan and Palestinian communities.

1.3.1 The Amish and Anabaptist communities

The Amish are a distinct group of traditional rural-living Anabaptist Christians, known for their simple living, plain dress and separation from modern technologies and conveniences, whose roots can be traced back to the Protestant Reformation in 16th century Europe. One group of reformers, the Anabaptists, rejected infant baptism, believing instead that only adults who had confessed their faith should be baptised (or

re-baptised). They also advocated for separation of church and state. Due to these beliefs, Anabaptist groups were severely persecuted throughout Europe, with many fleeing to the mountains of Switzerland and southern Germany.

One of the church leaders, a Swiss bishop named Jakob Ammann, sought to revitalise the Anabaptist movement by suggesting reforms of church practices that enforced closer links between church discipline and social practices. Amman proposed holding communion twice a year rather than once, as was the typical Swiss practice, and suggested that Christians, in obedience to Christ, should wash another's feet in the communion service. To promote doctrinal purity and spiritual discipline, Ammann forbade the trimming of beards and the wearing of fashionable dress, teaching that church members should dress in a uniform manner. He also believed in a stricter interpretation of the doctrine of shunning or "Meidung", a practice based on the New Testament command not to associate with a church member who was not repentful of his sinful conduct. Ammann taught that if a member of the church was excommunicated because of an unrepented sin, they should be completely shunned or avoided by all church members, including refusal of goods from the offending individual or refusing to share a meal with them. This practice was not intended to be a punishment but was instead meant as a lesson to help the individual realise the error of his ways and to seek forgiveness.

These beliefs however caused a schism between Ammann's followers and the other Anabaptists in Switzerland and Alsace, leading to an eventual split from the group in 1693. Ammann's followers were first called the "Amman-ish" group, later becoming known as the Amish.

The Amish first began emigrating to North America in the early 18th century to escape persecution in Europe for their religious beliefs. The Amish had largely been landless tenant farmers in Europe, so the chance to own their own land was also an attractive proposition. They first settled near Lancaster County in Pennsylvania ("Penn's woods"), attracted by the promise of religious freedom as part of William Penn's "Holy Experiment" of religious tolerance, where a large community remains today. The first wave of Amish migration to the state was between 1736 and 1770 and comprised around 500 individuals.

A second wave of Amish immigrants estimated to be around 3000 in number, occurred between 1815-1860. Compulsory military service was becoming commonplace in Europe, and this combined with the economic opportunities in North America, was a strong driving force to leave Europe and seek a new life in the “New World”. These Amish immigrants settled in new areas outside Pennsylvania, including Ohio, Indiana, Iowa and Illinois. The last Amish settlement in Europe disappeared in 1937; none remain in Europe today.

Today, the Amish are found in 31 states in North America, the Canadian provinces of Manitoba, New Brunswick, Ontario, and Prince Edward Island, and the South American countries of Argentina and Bolivia. Nearly two-thirds still live in the three states where the first Amish settlements were established; Pennsylvania, Ohio and Indiana (40). All Amish people share common beliefs and practices, including adult baptism, non-violence, lay ministers, small local congregations with services held in the homes of members with no formal religious buildings, rural living and separation from the outside world, education only up to eighth grade, church-regulated and traditional plain style of dress, selective use of technology, use of horse-drawn transportation and speaking a German-derived dialect (41). There are however around 40 different Amish affiliations, each comprising groups of church districts united by social and religious practices, differing with regards to types of dress, rules about technology and restrictions on participating in American society.

Founder effect and inherited diseases in the Amish

There is unfortunately a higher incidence of specific autosomal recessive genetic disorders, such as hereditary spastic paraplegia (42) and Ellis-van Creveld syndrome (a form of dwarfism) (43), among the Amish compared to the general population due to a phenomenon known as the “founder effect”. This refers to a loss of genetic variation that occurs when a new population is established by a very small number of individuals, or founders, from a larger population (Figure 1.5). Rare, recessive disease-causing mutations, if present in unaffected carrier, may then accumulate and become enriched in the population as it expands over subsequent generations,

leading to an increased frequency of certain otherwise rare autosomal recessive conditions within these communities (44).

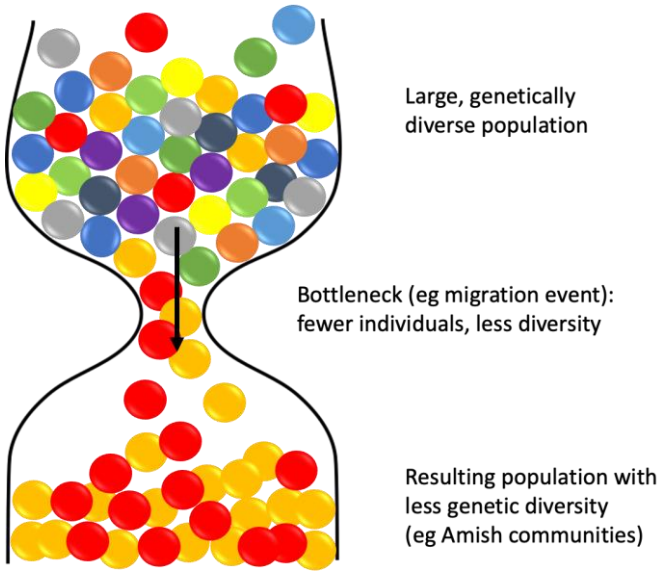


Figure 1.5 Ancestral bottleneck leading to founder effect

The modern Amish community in America is descended from the few hundred German-Swiss settlers who immigrated to the United States in the 18th century, many of whom were already part of related family groups. This limited number of ancestors founded a rapidly expanding Amish population, now estimated to number >350,000 in North America (40). One of the driving forces behind this population growth is attributed to large family sizes, with an average of 7-9 children per family. The Amish tend to have limited geographical mobility due to their religious constraints on transportation, and they tend to observe strict endogamy, marrying only within the community, with little or no migration into the community. This geographical and sociological isolation further contributes to the unique genetic distinctiveness of the Amish, providing a unique opportunity for the study of rare inherited disorders and identification of causative disease genes (45).

Genetic studies in the Amish are also aided by cultural and demographic characteristics of the community. The Amish keep extensive historical and genealogical records, enabling the construction of comprehensive family pedigrees. Large nuclear families (with low rates of non-paternity) are common, with a number of

unaffected as well as affected siblings with the same condition within each family, facilitating genetic analyses for identification of the causal disease variant. The impact of environmental factors on the expression of genetic diseases is minimised by the high standards of living, homogeneous lifestyles, and uniform socio-economic and occupational circumstances. Standards of medical care are relatively high, with extensive availability of medical records. The Amish also show an interest in illness and in the health of their relatives, with extensive sharing of health issues both locally and through publications such as “The Budget” newspaper, creating networks of knowledge about individuals with health problems. They also frequently provide home-based care for children with birth defects or genetic disorders instead of institutionalisation (46-48). Whilst studies of rare inherited diseases in modern diverse populations are hampered by small numbers of affected individuals and genetic and environmental complexities, these factors enhance the visibility of genetic disorders within the Amish communities, greatly facilitating the discovery of genes responsible for these inherited conditions.

Studies of inherited disorders present in the Amish were first undertaken in 1962 by the world-renowned geneticist Victor McKusick. These studies were inspired by an article written by a local family doctor, David Krusen, mentioning the frequent occurrence of achondroplasia among the Amish, as well as a manuscript entitled “Amish Society” by John Hostetler, which highlighted several characteristics of the Amish community that were advantageous to the study of genetic traits (47). The early studies on dwarfism in the Amish led to the recognition that the majority did not in fact have achondroplasia, and this instead led to the description of two separate forms of autosomal recessive dwarfism within the community; Ellis-van Creveld syndrome, a previously recognised disorder, as well as cartilage-hair hypoplasia, a new form of metaphyseal chondrodysplasia (47).

Over the next few decades, McKusick and his colleagues published over 30 reports of genetic disorders among the Amish (49). The work of McKusick led to the development and publication of the *Mendelian inheritance in Man* (MIM), first published in print as a trilogy of catalogues containing overviews of genes and genetic phenotypes for autosomal dominant, autosomal recessive and X-linked conditions (50). Print editions have since been superseded by an online version, the Online Mendelian of Inheritance

in Man or OMIM, providing a comprehensive and authoritative compendium of human genes and genetic phenotypes that is updated daily, and is a frequently used and invaluable resource in the field of medical genetics (51).

The Amish are surprisingly open to participation in genetic studies, and this receptivity may stem from the initial approaches by physicians and geneticists who were genuinely interested in helping affected individuals in the community. For instance, in patients with Ellis-van Creveld syndrome, arrangements were made for surgical repair of the associated cardiac defects and for orthopaedic correction of knee deformities. Researchers working with the community have also reported a willingness for families to participate in studies, even when their affected children may not necessarily benefit, because of the potential benefit to others, reflecting their generous and altruistic nature.

The higher incidence of rare inherited disorders among the Amish places a significant social and financial burden on this community. Whilst community schemes are available to help with medical expenses, the rising costs of medical care means that many families experience difficulties in obtaining clinical investigations and treatments. The Windows of Hope (WoH) Project was established in 2000, and is a non-profit translational community genomics research program led by Professor Andrew Crosby and Dr Emma Baple (research supervisors). It aims to determine the clinical and molecular spectrum of inherited diseases within the Amish and other Anabaptist communities by facilitating collaborations between clinical specialities as well as research and diagnostic genomic laboratories. The project has assisted in the identification and description of the underlying genetic cause in a large number of inherited disorders, including 18 novel conditions amongst the Amish and Anabaptist communities. These findings have allowed numerous families to receive much-needed diagnoses for previously unrecognised conditions, leading to an improved understanding of the clinical course and associated features, thereby facilitating improved clinical and developmental outcomes for the affected individuals through early medical intervention and social support.

To date, researchers have described over 250 separate genetic disorders among the Amish and other Anabaptist communities, collated into several community-specific

databases (52) (WoH Project Database, viewed at https://wohproject.com/). These include conditions that are completely new to medical science, and although a few disorders are unique to the Amish so far, most disease genes initially identified in the Amish have subsequently been shown to cause similar diseases in other populations worldwide, highlighting the global importance and relevance of studying inherited conditions in genetic isolates such as the Amish. There are also rare genetic disorders described in very small numbers of affected individuals identified in modern diverse populations, but present at increased frequency in the Amish communities due to an enrichment of disease-causing mutations. The larger numbers of affected individuals within the Amish communities facilitates the study of these rare genetic disorders, and may illuminate otherwise obscure biochemical pathways and enable a more comprehensive understanding of physiologic processes in human health and disease. This enhanced knowledge of the genetic and molecular mechanisms underpinning human disease may in turn lead to the development of new treatment strategies, thereby benefiting affected individuals worldwide. The increased numbers of affected individuals with a disease in a community also facilitates clinical studies to determine the clinical minutia of a condition, and to assess the efficacy of potential treatments as they are developed.

1.3.2 Inherited diseases in Pakistan

Pakistan is the 5th most populous country in the world, with a population exceeding 216 million (53), and is comprised of four provinces (Punjab, Khyber-Pakhtunkhuwa or KPK, Sindh and Balochistan) and three territories (Islamabad Capital Territory, Gilgit-Baltistan and Azad Kashmir). The population of Pakistan is ethnically and linguistically diverse, with around 18 ethnicities (based on historical lineage, geographical origin, language and cultural practices) and over 60 spoken dialects (54). The major ethnic groups are Punjabi and Pashtun (also known as Pathan) in northern Pakistan and Sindhi, Saraiki, Muhajirs (the immigrants from India, mostly Urdu-speaking) and Baloch in southern Pakistan (55). Social stratification is strongly based on ethnic and tribal groupings, with tribal groupings based on the *biraderi* or 'brotherhood' system, a patrilineal kinship that transcends geographical boundaries and is determined by family background and occupation (56).

There is a strong preference for endogamous marriage in Pakistan, where it is estimated that 40-60% of all marriages occur among relatives, particularly between first and second cousins (54, 55, 57). This arises in part due to the clan-orientated nature of the society, which values similarities in social group identity based on religion, ethnicity and tribal/clan affiliation (54). Cultural preferences for consanguineous marriages include the belief that the marriage will be more stable and secure with high fertility and lower divorce rates. There might already be existing family relationships between partners who are likely to have similar socioeconomic status and family customs, and there are financial benefits such as strengthening of the family business and maintenance of property within the family (55, 58). Most marriages still occur within the *biraderi*, with inter-*biraderi* marriages rare due to social stratification and limited social mixing (56). The frequency of autosomal recessive disorders is higher in populations with a high degree of endogamy, with Pakistan having one of the highest rates of genetic disorders in the world (59). The high rates of traditional intra-familial marriages, strong socio-cultural and ethnic divides, and geographical barriers and isolation has led to numerous genetically distinct communities, where there may be a high prevalence of specific inherited disorders due to the accumulation of regional founder variants (60, 61). These founder mutations often represent important causes of disease in a particular region, and knowledge of their presence, frequency and clinical outcomes will lead to enhanced knowledge of the genomic architecture of inherited diseases in Pakistan. This knowledge is of enormous value for local healthcare resource planning and will facilitate the design of community-specific hierarchical strategies for cost-effective genetic testing arrays to enable an accurate disease diagnosis to be achieved more rapidly, thus aiding the development of diagnostic and clinical care pathways and policies throughout Pakistan (59). This has led to the early development of specific genomic databases correlating disease mutation with clinical phenotype, geographic localisation and ethnicity or tribal affiliation within Pakistan (57).

1.3.3 Inherited diseases in Palestine

The Palestinians are a group of people who either live in, or originate from, historical Palestine, which formed the land bridge between the continents of Europe, Asia and Africa. Control of the region is strategic, and it remains one of the most contested

regions of the world today. Creation of the state of Israel in 1948 resulted in a displacement of a large proportion of the Palestinian Arab population, mainly to the West Bank and Gaza, but also to neighbouring Arab countries including Jordan, Lebanon and Syria. The Oslo accord of 1993 subsequently resulted in the establishment of Palestinian autonomy in the Occupied Palestinian Territories of the West Bank and Gaza. The global Palestinian population is now estimated to be 13.4 million, with the Palestinian diaspora numbering 6.7 million; 5.0 million in the Occupied Palestinian Territories, and 1.6 million in Israel (62). The majority of Palestinian Arabs are Muslims (over 80%); other religious communities are smaller and include Christians and the Duze (63).

Most of the Palestinian Arabs in Israel and in the territories under the Palestinian Authority live in villages or tribes that were founded by only a few individuals. With many Palestinians having large families (in 1993, the average number of offspring per woman in the West Bank and Gaza was 7.8) (64), there has been a natural expansion of the Arab population. Consanguineous marriages are common among the Palestinian Arab population, with over 40% of marriages occurring between relatives, and of these approximately half are between first cousins, most often involving children of brothers (65). Middle Eastern societies, and in particular the rural Arab populations, are characterised by close family relationships, and the preference for marrying relatives is a deeply rooted cultural trait. Although there appears to be evidence for a declining trend in consanguineous marriages amongst Palestinian Arabs (66-68), they remain prevalent within this population due to a number of political, economic and social factors that favour such marriages. Each village or tribe can therefore be considered a genetic isolate, where autosomal recessive diseases are relatively frequent, as the introduction of a new recessive mutation can lead to a relatively rapid appearance of several affected individuals in only a few generations. Whilst some inherited disorders such as thalassaemia and familial Mediterranean fever are prevalent throughout the Palestinian Arab population (69), the distribution of autosomal recessive diseases in general is not uniform, with many rare inherited diseases, including some novel diseases, specific to a single tribe or village due to this local founder effect (63). Inherited diseases are an important cause of childhood morbidity and mortality in these communities, however, particularly in the rural Palestinian areas, families often do not have access to advanced medical diagnostics

or assessment by trained specialists, and an accurate clinical diagnosis is often difficult to establish. A detailed knowledge of the distribution of genetic diseases in each village can therefore enable appropriate medical management as well as genetic counselling to be provided within the community (69).

1.4 Inherited eye diseases

Inherited eye diseases refer to a clinically and genetically heterogeneous group of disorders that can affect the structure and/or function of one or several structures of the eye, such as the cornea, anterior segment, lens or retina.

A large number of inherited eye disorders have been identified; these include conditions limited to the eye, as well as complex disorders such as chromosomal abnormalities or genetic syndromes with ocular manifestations that are associated with other systemic anomalies such as hearing impairment, learning disability or neurological deficits. The eye is an organ that is easily accessible for examination, and can provide a diagnostic window for wider systemic diseases; examples include the cherry red spot seen in Tay-Sachs and Sandhoff disease, angioid streaks in pseudoxanthoma elasticum, the iridescent “Christmas tree” cataract in myotonic dystrophy, lens subluxation in Marfan syndrome and homocystinuria, or the whorl-like corneal opacities or corneal verticillata seen in Fabry disease. Whilst inherited eye diseases are rare individually, together they contribute significantly to the burden of visual impairment and blindness.

Visual impairment is a major public health problem worldwide, with the World Health Organisation (WHO) estimating that in 2010, there were 285 million people globally of all ages who were visually impaired, of whom 39 million were blind (70) [according to the International Classification of Diseases Update and Revision 2006, visual impairment is classified as a visual acuity of less than 6/18, whilst a visual acuity of less than 3/60 would be classified as blindness (Snellen values) (71)]. In response to this global need, the WHO, in collaboration with the International Agency for the Prevention of Blindness (IAPB), launched the VISION 2020: The Right to Sight global initiative, aiming to eliminate avoidable blindness in the world by the year 2020, with the control of blindness in children (under age 16) a high priority (72, 73).

Inherited eye diseases contribute significantly to the burden of childhood blindness, accounting for approximately 40-50% of all childhood blindness in industrialised countries and 20-30% in developing nations (although this likely represents an underestimate due to poor clinical facilities and access in some regions) (73). Globally, the main inherited eye diseases that cause visual impairment are retinal dystrophies, corneal dystrophies, congenital and juvenile cataracts, aniridia and albinism (74). In the UK, inherited eye diseases account for a third of all children with severe visual impairment or blindness, with retinal dystrophies, microphthalmia/anophthalmia, nystagmus and albinism being important hereditary causes (75, 76). Hereditary retinal disorders are the most common cause of certifiable visual impairment in working age adults (accounting for a fifth of certifications in this age group) (77), and the third most common cause of certifiable visual impairment overall in the UK (78).

Blindness is a major disability that impacts on the physical, mental and social health of affected individuals, and places a significant emotional and financial burden on affected families and the wider community. As individuals with isolated inherited eye diseases often develop visual impairment at a young age and have a normal life expectancy, these conditions are associated with significant morbidity and socio-economic impact. Childhood blindness is a particularly significant problem in developing countries because many families to not have adequate access to healthcare and ophthalmic services due to a lack of infrastructure, resources and funding in some regions.

Establishing a precise clinical and molecular diagnosis in individuals with inherited eye diseases usually involves a thorough clinical examination to delineate the ocular phenotype and establish the presence or absence of any associated systemic features, a genetic workup including pedigree construction to show medical histories and genetic relationships within a family that might suggest a possible mode of inheritance, followed by chromosomal or molecular analysis to try identify the underlying genetic aetiology. An accurate disease diagnosis can bring about many important benefits for both patients and their families. It improves understanding of the disease which aids clinical management decisions and provides useful prognostic information for affected families. It can highlight the need for periodic screening to

enable early diagnosis or facilitate early treatment to prevent visual morbidity, such as prophylactic laser or cryotherapy treatment to prevent retinal detachment in individuals with Stickler syndrome (79). An accurate molecular diagnosis enables informed genetic counselling about the medical implications for the affected individual and members of the wider family, and it may provide advice on recurrence risk crucial for couples who wish to have more children. Determination of carrier status may provide reassurance that the risk of having affected children is low, or it may highlight a need for prenatal diagnostic evaluation and follow-up.

Determining the underlying genetic diagnosis in inherited eye diseases however is often challenging due to the phenotypic and genetic heterogeneity of these conditions. For instance, pleiotropic disorders such as Bardet-Biedl syndrome (BBS) can create diagnostic difficulties due to the multiple, seemingly unrelated phenotypic manifestations of the disease (80). Some inherited eye diseases such as inherited retinal dystrophies demonstrate remarkable genetic heterogeneity with overlapping clinical phenotypes; to date there are over 271 disease-associated genes (RetNet https://sph.uth.edu/retnet/, accessed 28.04.2021), spanning the entire spectrum of Mendelian (autosomal dominant, autosomal recessive and X-linked) and non-Mendelian (mitochondrial and digenic) inheritance (Figure 1.6). Recent advances in molecular diagnosis have also highlighted numerous inaccuracies in the traditional phenotype-based corneal dystrophy classification due to the genetic and phenotypic heterogeneity associated with the condition. For instance, mutations in two separate genes (*KRT3* and *KRT12*) have both been shown to cause the Meesman dystrophy phenotype, whilst several distinct corneal dystrophy phenotypes including Reis-Buckler, Thiel-Behnke, granular, lattice and Avellino corneal dystrophies have now all been shown to be caused by mutations in the same *TGFBI* gene. This has led to a revised classification by the International Committee for Classification of Corneal Dystrophies (IC3D) incorporating this new genetic understanding of corneal dystrophies, with *TGFBI*-associated dystrophies now classified into a single grouping (81, 82).

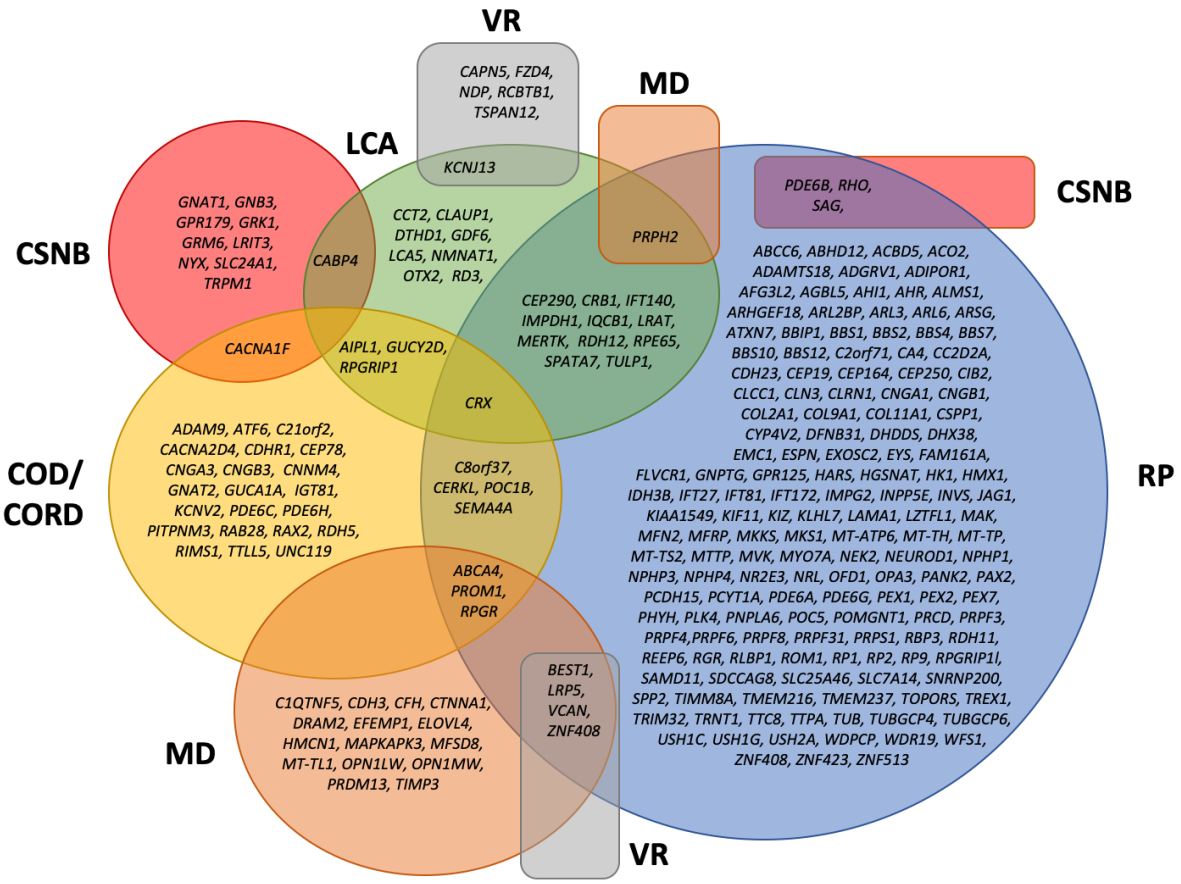


Figure 1.6 Genes associated with inherited retinal dystrophies

The coloured circles represent clinical diagnoses; overlapping circles indicate the significant clinical and genetic heterogeneity associated with inherited retinal dystrophies. Abbreviations: COD, cone dystrophy; CORD, cone-rod dystrophy, CSNB, congenital stationary night blindness, LCA, Leber congenital amaurosis; MD, macular dystrophy, RP, retinitis pigmentosa; VR, vitreoretinopathy. Modified from Berger et al (83)

Establishing an accurate clinical and molecular diagnosis is a particularly significant problem in developing countries, where there is limited access to ophthalmic equipment and expertise needed for performing and interpreting the detailed phenotyping studies needed for diagnosis due to regional geographical limitations or restricted local clinical resources. This problem is further compounded by a relative lack of knowledge regarding the specific nature and causes of inherited ocular diseases in developing nations.

Advances in next generation sequencing (NGS) technologies have ushered in a new era in molecular diagnostics by facilitating the sequencing of whole exomes or genomes of large cohorts of individuals with inherited eye diseases. This has not only

led to the identification of numerous new genetic causes responsible for a wide spectrum of inherited eye diseases such as corneal dystrophies, retinal dystrophies, and ocular developmental defects (84, 85), but may also permit a hypothesis-free diagnostic approach in settings where clinical information is limited, a situation commonly encountered in developing nations due to geographical isolation and limited access to healthcare services (see thesis chapter 5.3). There is a global responsibility to support genomic health initiatives in these nations so that the transformative potential of genomic medicine can be translated into meaningful health outcomes that benefit populations worldwide (59, 86).

1.4.1 Gene therapy in inherited eye diseases

Inherited eye diseases have historically played an important role in defining the basic principles of genetics. In 1876, the Swiss ophthalmologist Johann Friedrich Horner described the characteristic X-linked recessive pedigree pattern in a family with colour blindness, furthering progress in the understanding of sex-linked heredity; Horner's law stated that "colour-blind fathers have colour-normal daughters; and these colour-normal daughters are the mothers of colour-blind sons" (87). In 1936, Imai and Moriwaki first suggested cytoplasmic inheritance (now known as mitochondrial inheritance) to account for the maternal inheritance pattern of Leber hereditary optic neuropathy (88). The patchy retinal pigmentary pattern seen in heterozygote carriers of X-linked ocular albinism was seen to support the Lyon hypothesis, which postulated the random inactivation of one of the X chromosomes in somatic cells of females (89). Knudson's observations on retinoblastoma led to the development of the two-hit hypothesis in 1971, fundamental to our understanding of tumour suppressor genes and familial cancer syndromes (90). And the first convincing report of digenic inheritance in human disease was in 1994 for retinitis pigmentosa caused by mutations in the *PRPH2* and *ROM1* genes (91).

Today, advances in molecular genetics have in turn contributed significantly to ophthalmic genetics, advancing our understanding of the molecular biology of eye development and vision and the pathophysiological basis of eye diseases. For instance, mechanistic dissection of photoreceptor degeneration in retinal dystrophies identifies genes affecting a wide variety of cellular functions, including ciliary transport,

phototransduction and the visual cycle, lipid oxidation, protein degradation, cell signalling and cell-cell interactions, intracellular transport, phagocytosis and RNA splicing (92). Knowledge of the underlying genetic and molecular causes of disease can allow redeployment of existing therapies in targeting the biochemical pathways involved in disease pathogenesis, which may slow or halt disease progression, or enable restoration of protein or cell function through gene therapy or other experimental approaches.

The eye has a number of advantages as a target organ for gene therapy (93, 94). It is easily accessible by surgical approaches, and ocular tissues can be imaged and quantified *in vivo* with non-invasive imaging techniques such as electroretinography. Due to its small anatomical size and subdivision into smaller compartments, only small volumes of gene delivery vectors need to be administered to ensure delivery to target tissues. The relatively tight blood-retina barrier allows concentration of vectors within the target area and limits potential complications from systemic exposure, whilst at the same time altering the trafficking of immune cells from the systemic circulation to the eye and providing an “immune-privileged” state that limits the immune reaction to a given gene vector. The duplicity of the eye as an organ, and the bilateral nature of many inherited eye diseases, allows the option of treating just one eye, and considering the untreated eye as an ideal within-subject experimental control for comparisons on the safety and efficacy of treatment.

Gene therapy approaches have mainly focused on adeno-associated virus (AAV) or lentivirus vector-based gene replacement strategies in inherited retinal dystrophies, conditions for which management options were previously limited to visual rehabilitation, educational and social support. The recent approval of Luxturna® (voretigene neparvovec) for the treatment of *RPE65*-mediated inherited retinal dystrophy in Europe and the United States (95, 96), together with early promising results from clinical trials currently underway for a further range of retinal dystrophies associated with *CHM*, *PDE6B*, *RPGR*, *MERTK*, *MYO7A*, *ABCA4*, *RS1*, *CNGA3* and *CNGB3* mutations (94), marks a new era in the treatment of these previously incurable and sight depriving conditions. Non-viral systems for gene delivery such as electroporation, nanoparticles and liposomes are also being studied, and although

these will be cheaper and easier to produce, they currently do not seem to show the same therapeutic promise (94).

Antisense oligonucleotides are another promising therapeutic strategy that can be used to target disease mutations, with several different approaches currently undergoing clinical trials in inherited retinal dystrophies. One approach uses the antisense molecule to block the activation of cryptic splice sites and restore normal splicing mechanisms, and is useful for diseases resulting from aberrant splicing; this approach is currently under investigation for the treatment of Leber congenital amaurosis due to a splice site mutation in the *CEP290* gene (97). Another approach based on splice modulation is being studied for the treatment of syndromic and non-syndromic retinitis pigmentosa caused by mutations in the *USH2A* gene; a common site of disease mutations is within exon 13 of this gene, and the antisense molecule is designed to cause skipping of this exon from the *USH2A* mRNA, resulting in an in-frame transcript that is translated into a shorter but still functional usherin protein (98). A further approach can be used for diseases caused by dominant negative mutations, where the antisense molecule targets and degrades the mutant mRNA; this approach is currently being trialled for the treatment of autosomal dominant retinitis pigmentosa caused by *RHO* gene mutation (99).

Another gene-based therapeutic strategy under investigation is the use of small molecule translational read-through inducing drugs or TRIDs such as PTC124 (Ataluren) or aminoglycoside antibiotics such as gentamicin (G418). These act by interfering with the proof-reading abilities of translationally active ribosomes, allowing “read-through” of premature termination codons (PTCs) and translation of a full-length protein from the mutant mRNA. This approach may be useful for recessive diseases where nonsense mutations are a significant cause of disease, and where a small amount of functional protein may be sufficient to rescue the disease phenotype, and has been studied for a number of inherited ocular conditions including choroideraemia, retinitis pigmentosa, Usher syndrome and aniridia (100). TRIDs have several advantages compared to gene replacement strategies; the size of the gene is not crucial (compared to the limited packaging capacity of commonly used viral vectors), and the targeted genes remain under endogenous control, maintaining tissue-specific

timing and duration of gene expression and splicing. Concerns remain however over long-term toxicity and the need for repeated dosing.

Recently, the first in-human trial of genome editing using the clustered regularly interspersed palindromic repeat (CRISPR)/Cas9 system has been performed to target and delete a cryptic splice site and restore normal splicing in individuals with Leber congenital amaurosis (101), with plans for a similar trial targeted to Usher syndrome (102, 103).

Although progress is not as advanced as for inherited retinal dystrophies, gene therapy approaches are also being investigated as a therapeutic option for other inherited eye diseases including Leber hereditary optic neuropathy (104), mucopolysaccharidoses (105), Fuchs corneal endothelial dystrophy associated with expansion of a non-coding trinucleotide repeat in *TCF4* (106), and *MYOC*-associated glaucoma (107).

Challenges remain in the development of molecular therapies for inherited eye diseases, both in terms of technical hurdles, and the social responsibility to provide equitable access to these transformative therapies given the significant human cost of sight loss for affected individuals (108). Despite this, advances in molecular technologies have greatly expanded the diagnostic capabilities and therapeutic landscape for inherited eye diseases, providing an optimistic outlook for individuals with these previously untreatable conditions.

1.5 Disease gene and variant identification strategies

Identification of disease genes enables an accurate molecular diagnosis in affected individuals, resulting in improved clinical care. It is also crucial to understanding the underlying genetic and molecular pathways of disease, which may in turn highlight potential therapeutic targets. The rapid evolution and widespread introduction of next generation sequencing (NGS) technologies such as whole exome and whole genome sequencing in research as well as clinical diagnostic settings has allowed rapid progress in the identification of genes and variants associated with inherited eye diseases. The combined approach of NGS in parallel with autozygosity mapping, a powerful tool in the study of autosomal recessive diseases in endogamous

communities, has been successful in uncovering novel causes of inherited eye diseases in community settings (38, 60), and was the primary approach adopted in many studies outlined in this thesis. Further background information on autozygosity mapping and advances in NGS technologies is contained in Appendix A

1.6 Project aims

The overarching objective of this PhD project involves the delineation and characterisation of the phenotypic, genetic and molecular spectrum of inherited eye diseases in community settings including the Amish, Pakistani and Palestinian communities. This will increase scientific understanding of the genetic causes and molecular pathways underlying ocular diseases, and drive advances that will benefit patients worldwide with the same condition. In order to achieve this, the specific aims of the study include:

1. screening of established disease genes in affected families with isolated and syndromic inherited eye diseases including: anophthalmia, infantile nystagmus and oculocutaneous albinism (OCA), anterior segment dysgenesis and congenital cataracts, as well as inherited retinal dystrophies to identify previously reported and novel pathogenic gene variants. Alongside this, comprehensive literature reviews will be undertaken to determine the clinical and molecular spectrum of inherited eye diseases in communities, facilitating the development of targeted genetic testing strategies that will enable improved and efficient disease diagnosis and early intervention,

2. investigation of the pathogenicity of two common *TYR* gene variants through co-segregation studies in Amish OCA families, together with analysis of extensive UK and international OCA patient cohorts and functional assays. This will provide novel insights into OCA molecular pathogenesis, and help resolve a long dispute regarding the pathogenicity and clinical relevance of these gene variants,

3. perform detailed clinical phenotyping in combination with comprehensive genomic studies in molecularly undiagnosed individuals with inherited eye diseases. This will enable the identification, consolidation and characterisation of ultra-rare causes of inherited eye diseases in Amish, Pakistani and Palestinian communities.

2 MATERIALS AND METHODS

2.1 Materials

General laboratory consumables were purchased from Rainin, STARLAB and Alpha Laboratories. Reagents used are reported in Table 2.1.

Table 2.1 Reagents used in the study

Reagents	Manufacturer
Agarose (molecular grade)	Fisher Scientific
Boric acid	Fisher Scientific
dNTP set	Solis BioDyne
DMSO	Sigma-Aldrich
DreamTaq™ DNA polymerase	Thermo Fisher Scientific
DreamTaq™ green buffer (10x)	Thermo Fisher Scientific
Ethanol	Fisher Scientific
Ethidium bromide	Fisher Scientific
Exonuclease I (Exo I)	New England Biolabs
GeneRuler™ 1kb Plus DNA ladder	Thermo Fisher Scientific
Lithium acetate dihydrate	Alfa Aesar
Shrimp alkaline shosphatase (rSAP)	New England Biolabs

2.2 Clinical methods

2.2.1 Ethical approval for study

Ethical approval was granted from the ethical approval committees of:

- University of Exeter Medical School, Exeter, UK
- South Central – Hampshire A Research Ethics Committee, Southampton, UK
- Akron Children's Hospital, Ohio, USA
- Baylor College of Medicine, Texas, USA
- University of Arizona, Arizona, USA
- International Islamic University, Islamabad, Pakistan
- Kohat University of Science and Technology, KPK, Pakistan
- University of Health Sciences, Lahore, Pakistan

- Shah Abdul Latif University, Sindh, Pakistan
- Arab American University, Jenin, Palestine

The study protocol adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to their inclusion in this study with parental written consent provided on behalf of children involved in the study.

2.2.2 Patient ascertainment and phenotyping

Recruitment of families from the North American Amish communities, as well as communities in Pakistan and Palestine, was performed by collaborating researchers. All study participants were phenotypically assessed and overseen by local clinicians. A medical history was obtained, including a documentation of visual symptoms. Facial photographs and videos were used to document skin and hair tone and nystagmus. Visual acuity testing was performed using Snellen charts, and colour vision testing using Ishihara charts. The anterior and posterior segment of the eye was examined by slit lamp biomicroscopy or direct fundoscopy. Additional investigations were performed in selected affected individuals where possible; these include retinal imaging studies such as colour fundus photography and ocular coherence tomography (OCT), as well as electrophysiological testing including electroretinography. Subsequent to molecular genetic studies, additional targeted clinical investigations were occasionally undertaken to clarify the clinical significance of any candidate variants identified.

Considerable diagnostic difficulties were encountered by clinicians in some communities, where affected families resided in rural regions with a lack of medical infrastructure, limiting access to specialised equipment required for detailed and accurate ocular phenotyping. Local clinicians and scientists were supported by myself (clinical ophthalmologist) in the collection and review of the highest quality clinical phenotypic data possible.

2.3 Molecular genetic methods

2.3.1 Sample acquisition and data management

The research study was carried out in compliance with the Human Tissue Authority (HTA) Codes of Practice and Standards (Code E: Research). All blood and buccal samples and subsequent DNA extractions used in this project were stored in HTA-licensed premises with research carried out in accordance with the Human Tissue Act 2004. Each participating individual was assigned a unique study identification number. A master list of study participants linking the identification number to the study individual was recorded in a password protected database and stored on university servers. All data management was GDPR compliant.

Participating individuals had either peripheral venous blood samples taken in EDTA-containing vacutainer tubes, or buccal cell collection using the ORAcollect® for paediatrics kit (DNA Genotek). Blood and buccal sample tubes were labelled with their study identification number on arrival; blood samples were stored at -20°C, whilst buccal samples were stored at room temperature, prior to DNA extraction.

2.3.2 DNA extraction and quantification

DNA extraction from whole blood

Genomic DNA extraction for venous blood samples was performed using the ReliaPrep™ Blood gDNA Miniprep System (Promega). Blood samples were thoroughly mixed by hand for at least 2 mins. 200 µl of blood was added to 20 µl of proteinase K solution in a 1.5 ml microcentrifuge tube and mixed briefly. 200 µl of cell lysis buffer was then added to the tube, which was capped and mixed by vortexing (Topmix FB15024 Vortex Mixer, Fisher Scientific) for at least 10 sec, followed by an incubation for 10 mins at 56°C on a heating block (Mixer HC Thermoblock, STARLAB). After this incubation period, the tube was removed from the heating block, and 250 µl of binding buffer was added to the tube, which was then capped and mixed for 10 sec using a vortex mixer. At this stage, the lysate was visually checked to ensure a dark green colour as per protocol specifications.

The contents of the tube were then added to a ReliaPrep™ binding column placed in an empty collection tube, which was capped and centrifuged for 1 min in a microcentrifuge (Biofuge Pico Microcentrifuge, Heraeus) at maximum speed (13,000 rpm, 16,060 g-force). The binding column was checked to make sure that the lysate had completely passed through the membrane, and if there was lysate still visible on top of the membrane, the column was centrifuged for a further 1 min. The collection tube containing the flow-through was then removed, and the liquid discarded as hazardous waste. The binding column was placed in a fresh collection tube, and 500 µl of column wash solution was added to the column, this was then centrifuged for 3 mins at maximum speed, and the flow-through discarded. This last step was repeated twice more for a total of three washes.

The column was then placed in a clean 1.5ml microcentrifuge tube, and 200 µl of nuclease free water was added to the column. This was centrifuged for 1 min at maximum speed to elute the DNA.

DNA extraction from buccal cells

Genomic DNA extraction for buccal samples was performed using the *Xtreme* DNA kit (Isohelix). The sample tubes containing the stabilised swab heads were first incubated in a water bath at 50°C for 1 hour. 20 µl of proteinase K solution was then added to the sample tube, which was immediately mixed by vortexing (Topmix FB15024 Vortex Mixer, Fisher Scientific), and then further incubated in a water bath (Unstirred Digital Water Bath, Clifton) at 60°C for 1 hour. 750 µl of column binding buffer was added to the tube, which was mixed by vortexing for 30 sec. 1.25ml of ethanol was then added to the tube, and this was mixed by vortexing.

700 µl of the sample was added to an *Xtreme* DNA column placed in a collection tube, and this was centrifuged for 1 min in a microcentrifuge (Biofuge Pico Microcentrifuge, Heraeus) at maximum speed (13,000 rpm, 16,060 g), with the flow-through then discarded. This step was repeated until all the sample had been loaded onto the column. 750 µl of wash buffer solution was then added to the column, and this was centrifuged at maximum speed, with the flow-through discarded. This step was

repeated once for a total of 2 washes. The column was then placed in a clean collection tube and centrifuged at maximum speed for 3 mins to remove all traces of ethanol.

Following this, the column was placed in a clean 1.5 ml microcentrifuge tube and 100 µl of elution buffer (preheated to 70°C) was added to the centre of the membrane. The column was allowed to stand for 3 mins, following this it was then centrifuged at maximum speed for 1 min to elute the DNA.

DNA quantification and storage

Extracted DNA was quantified using a spectrophotometer (NanoDrop™ 2000, Thermo Fisher Scientific) to measure the absorption of UV light at 260 nm (peak absorbance wavelength of nucleic acids) in a 1 µl sample aliquot. The NanoDrop™ software automatically calculates the nucleic acid concentration (in ng/µl) using a modified Beer-Lambert equation, which correlates the calculated absorbance with concentration.

The quality of the extracted DNA was also assessed simultaneously by measuring the absorption of UV light at 280 nm (strong absorbance wavelength of proteins and phenolic compounds) and 230 nm (strong absorbance wavelength of organic compounds). The absorbance at 260 nm, 280 nm and 230 nm normalised to a 10 mm pathlength is denoted by A260, A280 and A230 respectively, and the A260/A280 and A260/A230 ratios were used to assess DNA purity. A A260/A280 ratio of ~1.8 generally indicates “pure” DNA; an appreciably lower ratio indicates the possible presence of protein or other contaminants that absorb strongly at or near 280 nm. The A260/A230 ratio is a secondary measure of nucleic acid purity, with values for “pure” nucleic acid often higher than the respective A260/A280 values and commonly in the range of 1.8 - 2.2; an appreciably lower ratio may indicate the presence of co-purified contaminants. Following extraction and quantification, DNA was stored at 4°C.

2.3.3 Polymerase chain reaction (PCR) and dideoxy sequencing

Primer design

Oligonucleotide primer pairs for dideoxy sequencing were designed using reference genomic sequences accessed from the UCSC Genome Browser (https://genome.ucsc.edu/) and the open-source software Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) for primer design. Occasionally, it was necessary to design some primers by eye where the regions flanking the target sequence were highly repetitive or GC rich and hence not conducive to automatic primer design. Primers were selected based on the following principles:

- Primer size between 20-26 nucleotides in length
- Similar predicted melting temperatures for the primer pair
- Primers that specifically anneal only to the region of interest with no alternative binding sites in the genome identified [checked using the human BLAT search tool at UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgPcr)]
- Primers that did not contain any common single nucleotide polymorphisms (SNPs) (≥1% minor allele frequency or MAF in dbSNP build 151) or repetitive elements (including low-complexity sequences and interspersed repeats) identified using the RepeatMasker software tool [checked using the human BLAT search tool at UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgPcr)]
- A PCR product size of between 200 to 1000 bp in length (where possible)
- An PCR product GC content of less than 60% (where possible)

Oligonucleotide primers were ordered and manufactured by Integrated DNA Technologies IDT™ and were received in lyophilised form. Primers were re-suspended in double-distilled water (ddH_2O) as per manufacturer's instructions to achieve a 100 µM stock solution and stored at -20°C. A subsequent 5 µM working dilution of 200 µl was made for PCR amplification and sequencing, and was stored at 4°C. Specific primer details are contained within Appendix D.

PCR conditions

PCR reactions for each primer pair were first optimised using a temperature gradient protocol to determine the optimal primer annealing temperature for achieving a high yield of PCR product with minimal non-specific amplification (Figure 2.1). The temperature gradient would typically be run across 12 reactions, with each reaction having a different annealing temperature ranging from 52-64°C, each differing by approximately 1°C.

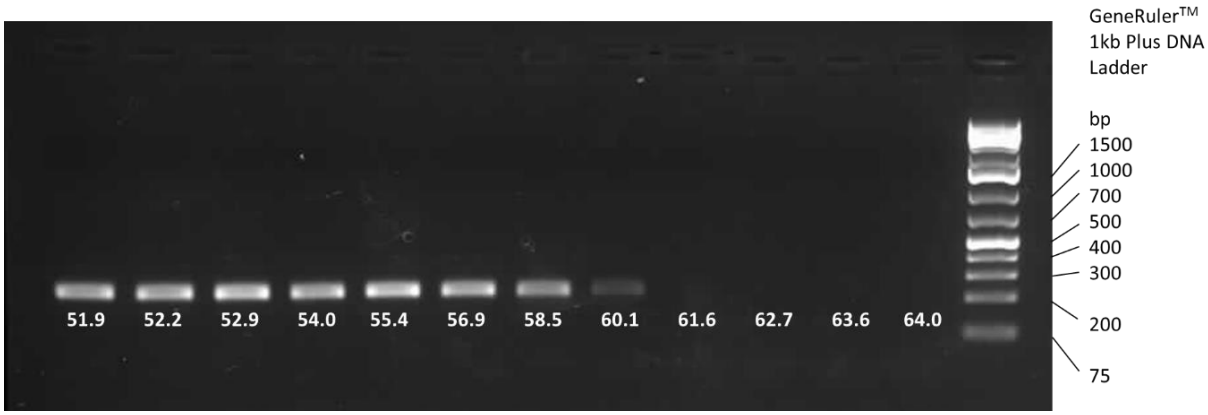


Figure 2.1 Temperature gradient optimisation example

Agarose gel electrophoresis image of PCR reaction using primers designed to amplify the OCA2 p.(Ala787Val) variant. 58°C was the optimal annealing temperature chosen as it was the lowest temperature with a bright and clear band.

PCR was performed using a 10 µl reaction mixture (Table 2.2). The DreamTaq™ Green Buffer contains $MgCl_2$ (at a concentration of 20 mM) and is optimised for PCR reaction with the DreamTaq™ DNA Polymerase, the standard Taq polymerase used in the lab. The buffer also contains a density reagent and two tracking dyes, allowing direct loading and visualisation of PCR products onto the agarose gel, as well as facilitating the monitoring of the electrophoresis progress. The dNTP Mix solution was prepared from stock solutions of 100mM dATP, dCTP, dGTP and dTTP (dNTP Set, Solis BioDyne).

Table 2.2 Standard PCR reaction mixture

Component (concentration)	Volume
Primer forward (5 µM)	0.4 µl
Primer reverse (5 µM)	0.4 µl
DreamTaq™ green buffer (10x)	1.0 µl
DreamTaq™ DNA polymerase (5 units/µl)	0.1 µl
dNTP mix solution (10 mM)	0.4 µl
DNA (10-30 ng/µl)	0.8 µl
ddH_2O	6.9 µl
Total	10 µl

Modifications to the standard PCR reaction mixture were occasionally required. For instance, where the PCR products had an unavoidably high GC content (typically >60%) and where there was non-specific priming or poor PCR product yield, the PCR reaction was supplemented with the addition of 10% DMSO. In DNA templates with a high GC content, there is increased difficulty in denaturing the template due to the increased hydrogen bond strength, causing intermolecular secondary structures to form more readily, which can compete with primer annealing (109). DMSO is thought to resolve secondary structure formation by binding to the major and minor grooves of the template DNA, hence destabilising the double helix structure and promoting denaturation (109).

PCR was performed in a thermal cycler (Mastercycler® ep gradient S, Eppendorf) following a touchdown PCR protocol (Table 2.3). This method aims to increase specificity of the PCR reaction by reducing off-target priming, and uses an initial annealing temperature (T_a) that is greater than the projected melting temperature of the primers, followed by a gradual decrease in annealing temperature over subsequent cycles until the desired annealing temperature (as identified through temperature gradient optimisation) is achieved (110).

Table 2.3 Touchdown PCR protocol

Step		Temperature	Time
1	Initial denaturation	95°C	5 mins
2	Denature	95°C	30 sec
3	Anneal	T_a + 4°C	30 sec
4	Extension	72°C	30-60s
Repeat steps 2-4 for a total of 2 cycles			
5	Denature	95°C	30 sec
6	Anneal	T_a + 2°C	30 sec
7	Extension	72°C	30-60 sec
Repeat steps 5-7 for a total of 2 cycles			
8	Denature	95°C	30 sec
9	Anneal	T_a	30 sec
10	Extension	72°C	30-60 sec
Repeat steps 8-10 for a total of 35 cycles			
11	Final extension	72°C	5 mins

Abbreviation: T_a, annealing temperature.

The duration of the extension step was modified depending on the size of the PCR product; for product sizes of less than 500 bp in length, an extension time of 30 sec was used, for PCR product sizes between 500 to 750 bp long, an extension time of 45 sec was used, and where the PCR product was over 750 bp in length, an extension time of 60 sec was used. Specific PCR conditions for each primer pair are contained within Appendix D.

Agarose gel electrophoresis

In agarose gel electrophoresis, the agarose gel forms a highly cross-linked matrix through which the negatively charged DNA molecule is forced to migrate in response to an electric current towards the positive anode. This technique can be used to separate DNA fragments by size, as shorter DNA molecules will migrate more rapidly through the matrix compared to longer molecules, and will travel a greater distance across the agarose gel in a given period of time. In this study, agarose gel electrophoresis was mainly used to determine the adequacy of the DNA amplification following PCR and to confirm the presence of a single amplicon.

1% agarose gels were made by heating 1.0 g molecular grade agarose powder with 100 ml 1X lithium acetate borate (LAB) buffer (consisting of 10 mM lithium acetate and 10 mM boric acid, made by dissolving 51 g of lithium acetate dehydrate and 31 g boric acid in 1 litre of distilled water to make a 1 litre 50X stock solution, before being diluted to 1X) in a 700 W microwave till the agarose has completely dissolved. The agarose solution was left to cool to approximately 50-60°C before 5 µl of 1% ethidium bromide (a DNA-binding fluorophore) was added, this was then poured into a gel tray with a well comb in place, and left to sit at room temperature for approximately 30 to 60 mins until the gel was completely solidified.

The agarose gel was then placed in an electrophoresis tank containing 1X LAB buffer, and 3 µl of PCR product was loaded directly into the wells (no loading buffer was required due to the use of DreamTaq™ Green Buffer for the PCR reaction), together with an appropriate molecular weight marker, most typically the GeneRuler™ 1kb Plus DNA Ladder, to allow sizing of the PCR product.

The gel was run at 150 V constant voltage for approximately 30 mins, and then imaged using an ultraviolet transilluminator, causing fluorescence of the ethidium bromide and allowing visualisation of the DNA banding pattern. The image was photographed using a gel imaging and analysis system (InGenius gel documentation system and GeneSnap image acquisition software, Syngene).

PCR clean-up and dideoxy sequencing

Prior to sequencing, unincorporated primers and nucleotides were removed using a combination of the hydrolytic enzymes exonuclease I (Exo I) (which degrades single stranded DNA including residual primers and superfluous PCR products) and shrimp alkaline phosphatase (rSAP) (which dephosphorylates and hence deactivates the remaining nucleotides) (111).

An ExoSAP enzyme mix was first prepared, consisting of 2.5 µl of Exo I (20,000 U/ml) and 25 µl rSAP (1000 U/ml), made up to 1 ml with ddH_2O. 2 µl of this ExoSAP enzyme mix was then added to 5 µl of PCR product, and this mixture was incubated at 37°C

for 30 mins (for digestion of excess primers and dephosphorylation of nucleotides), followed by a second incubation at 95°C for 5 mins for enzyme inactivation.

The ExoSAP treated PCR products were sequenced by Source BioScience (https://www.sourcebioscience.com/), with the generated chromatogram data file visualised using a chromatogram viewer (FinchTV, Geospiza).

2.3.4 Single nucleotide polymorphism (SNP) genotyping

Within community settings, affected individuals for autosomal recessive diseases are typically homozygous for founder mutations that occur at an increased frequency within such populations, and reside within identical-by-descent haplotypes (60, 112). Homozygosity mapping techniques, which highlight shared regions of homozygosity between affected individuals, therefore facilitate an efficient means of identifying candidate disease genes within these regions. In this study, this was performed using genome-wide SNP genotyping (by Dr Barry Chioza and Joe Leslie, University of Exeter) using the Illumina CytoSNP-12v2.1 array and following the Infinium® HD Assay Ultra manual protocol. This protocol is performed over 3 days with 200 ng of DNA at a concentration of 50 ng/µl per sample.

Day 1: DNA samples were denatured using a buffer containing 0.1 N sodium hydroxide, and then neutralised. Samples were amplified during an overnight incubation at 37°C.

Day 2: Amplified DNA samples were enzymatically fragmented using the Illumina fragmentation solution (FMS), using endpoint fragmentation to avoid overfragmentation. The DNA was then precipitated using 100% 2-propanol and the Illumina precipitation solution (PM1), and collected via a 20 min centrifugation performed at 4°C. The DNA was then resuspended using the Illumina resuspension, hybridisation and wash solution (RA1), and then denatured at 95°C for 20 mins. The denatured samples were cooled at room temperature for 30 mins before being loaded onto the BeadChips (with each chip holding 12 samples). This was then incubated in the Illumina hybridisation oven at 48°C for 16-24 hours.

Day 3: Beadchips were prepared for the staining process by washing with the Illumina Prepare BeadChip Buffer 1 (PB1) solution in order to remove unhybridised and nonspecifically hybridised DNA samples from the BeadChips. Labelled nucleotides were dispensed onto the BeadChip through the flow-through chambers to perform single-base extension of primers hybridised to the DNA samples. The BeadChips were then stained using the Illumina XStain HD BeadChip process and imaged on an Illumina iScan System. The iScan system uses a laser to excite the fluorophores of the single-base extension product on the beads of the BeadChip. Light emissions from the fluorophores were recorded by the reader, allowing high resolution images of the BeadChip to be taken. The data from these images was analysed using the Illumina GenomeStudio Integrated Informatics Platform, allowing the genotype to be determined.

Further analysis was undertaken using the Illumina KaryoStudio v1.4 software for identification of cytogenetic or chromosomal structural aberrations. The generated data was also exported into Microsoft Excel 2013, and analysed visually, as well as using an Excel macro written by Dr Barry Chioza, to highlight notable genomic regions (>1 Mb) with a shared haplotype.

2.3.5 Next generation sequencing (NGS)

TruSight™ One Sequencing Panel

The Illumina TruSight™ One clinical exome sequencing panel was performed by Luke O'Gorman and Chelsea Norman, University of Southampton. This panel provides targeted sequencing for 4813 genes associated with clinical phenotypes, including a number of known causative genes for inherited ocular disease. Next generation sequencing was performed using the Illumina TruSight™ capture kit and the NextSeq 500 platform (Illumina), with an approximate read depth and coverage of 20X minimum for > 95% of the genome.

Whole exome sequencing (WES)

WES was predominantly performed by either BGI Tech Solutions (BGI Genomics, Hong Kong) using a BGISEQ-500 platform or by the Exeter Sequencing Service (University of Exeter, UK) using a Illumina NextSeq 500 platform, with an approximate read depth and coverage of 20X minimum for 90-95% of the genome achieved with both platforms.

Bioinformatic pipeline

Alignment of NGS data to the reference human genome sequence

Raw sequence data in FASTQ format was aligned to the Genome Reference Consortium human genome build 37 (GRCh37) using the BWA-MEM (v0.7.12) (113) alignment algorithm, generating output files in Sequence Alignment/Map (SAM) format, which were then converted into Binary Alignment/Map (BAM) format using SamFormatConverter (Picard Tools v2.15.0) for further analysis. FixMateInformation (Picard Tools v2.15.0) was used to verify read pair information between each read and its corresponding mate pair and fix information errors if necessary, sorting the output file by read coordinates. Duplicate reads were removed using MarkDuplicates (Picard Tools v2.15.0). This identifies the duplicate reads by matching all read pairs with identical 5' coordinates and orientations, avoiding bias in variant calling by presenting variants with artificially high read depth support. Indel (insertion and deletion) realignment and base quality recalibration was performed using the Genome Analysis Toolkit (GATK) v3.70 (114).

Identification of sequence variation

Variant calling for SNPs and short indels was performed using GATK HaplotypeCaller (v3.70) with a variant call format (VCF) file generated for each individual. This was filtered for quality control based on read depth (DP), mapping quality (MQ), strand bias, and relative position of the variant within the read.

Copy number variants (CNV) (WES data only) were detected using ExomeDepth (115), which uses read depth data to call CNVs from exome sequencing experiments, and SavvyCNV (116), which calls CNVs using off-target read data.

Annotation of sequence variation

Functional annotation of variants in the filtered VCF file was then performed using Alamut® Batch (v1.10), facilitating interpretation of variants in a clinical context. Each variant was annotated on all available transcripts.

Variant prioritisation

Variants (SNPs and short indels) were prioritised using the following criteria (Table 2.4).

Table 2.4 Criteria used for variant prioritisation

Criteria	Description
Call quality	• VCF filter = PASS • MQ ≥ 50 • DP ≥ 5 Integrative Genome Viewer (IGV) was also used to visualise the aligned sequencing data and genomic region adjacent to the variant
Consistent with mode of inheritance	If pedigree suggests: • Autosomal recessive inheritance: homozygous and compound heterozygous variants • Autosomal dominant inheritance: heterozygous variants • X-linked inheritance: hemizygous variants (in males)
Frequency in control datasets	MAF <0.01 in: • gnomAD (v2.1.1 and v3.1.1) (all populations) • Relevant internal control exome dataset of unrelated Amish (220) or Pakistani (65) individuals
Predicted impact on protein or splicing	• Non-synonymous missense variants • Presumed loss of function variants • Splicing variants, defined as intronic variants within 5bp of the intron-exon junction
In silico pathogenicity predictions	Non-synonymous missense variants: • SIFT score < 0.05 • PolyPhen-2 score > 0.15 • PROVEAN score < -2.5 Splicing variants:

	• Local splicing effect predictions calculated using the Alamut® Batch interpretation algorithm that interprets splice site signals recognised by MaxEntScan, SpliceSiteFinder-like (SSF) and Splice Site Prediction by Neural Network (NNSplice)
Previously reported variants	• 'Pathogenic' and 'Likely pathogenic' annotation in ClinVar • 'Disease-causing mutation' and 'Disease-causing mutation?' annotation in HGMD professional v2020.4

For glossary of terms see Appendix A.

2.4 Literature review

Literature reviews were performed using Pubmed (https://pubmed.ncbi.nlm.nih.gov/), Google Scholar (https://scholar.google.com/) and HGMD® Professional 2020.4 (https://my.qiagendigitalinsights.com/bbp/view/hgmd/pro/start.php) to retrieve all reported disease-associated loci and variants. Additional information collated where relevant included the number of families and number of affected individuals within each family, their country of origin, and reported phenotype. The variants were then searched for within NCBI ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) to identify any additional literature reports and evidence supporting causality.

3 STUDIES OF OCULOCUTANEOUS ALBINISM (OCA) IN COMMUNITIES

3.1 Introduction

OCA refers to a group of genetically and clinically heterogeneous disorders characterised by abnormal melanin synthesis, resulting in decreased or absent pigmentation of eyes, skin and hair. Six genes involved in the melanin biosynthesis pathway are known to cause non-syndromic OCA, with gene variants in *TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *SLC24A5* and *C10orf11* associated with OCA subtypes 1, 2, 3, 4, 6 and 7 respectively. OCA5 has been mapped to the chromosome 4q24 locus, although the gene responsible has not yet been identified (117). OCA may also be a part of a broader systemic phenotype in conditions associated with lysosomal storage defects such as Hermansky-Pudlak syndrome (associated with pathogenic variants in *HPS1*, *HPS3*, *HPS4*, *HPS5*, *HPS6*, *AP3B1*, *DTNBP1*, *BLOC1S3*, *BLOC1S6* and *AP3D1*) and Chediak-Higashi syndrome (associated with pathogenic variants in *LYST*). Both syndromic and non-syndromic forms of OCA are inherited as autosomal recessive conditions.

Ocular features are present in individuals with OCA, and are characteristic of the disease. These include photophobia, nystagmus, foveal hypoplasia, iris transillumination and abnormal decussation of nerve fibres at the optic chiasm resulting in crossed asymmetry on visual evoked potential testing (118). These ocular features may however be variable with no single defining characteristic found to be present in every individual with OCA (119). The cutaneous phenotype may also vary, ranging from total absence to near normal levels of pigmentation, and can be difficult to evaluate, particularly in individuals with a lightly pigmented ethnic background (120, 121). As such, OCA can be difficult to distinguish clinically from several other ocular disorders with overlapping phenotypical features, such as *GPR143*-associated X-linked ocular albinism, where the hypopigmentation is limited to the eye (122), *FRMD7*-associated X-linked idiopathic congenital nystagmus (123), *SLC38A8*-associated foveal hypoplasia (also known as FHONDA; foveal hypoplasia, optic nerve

decussation defects, and anterior segment dysgenesis) (124), and dominant *PAX6*-related ocular developmental disorders (125).

OCA1, associated with *TYR* gene variants, is the most common OCA subtype found in Caucasians accounting for ~50% of cases worldwide (126, 127). *TYR* encodes the enzyme tyrosinase, which is the critical and rate limiting enzyme in the biosynthesis of melanin in follicular and epidermal melanocytes in hair and skin, as well as in uveal melanocytes in the iris, ciliary body and choroid, and RPE cells in the eye (128). Disease-associated variants in the *TYR* gene cause complete or partial OCA1 depending on their impact on the residual activity of the encoded mutant tyrosinase enzyme (129). *TYR* gene variants that result in a severe reduction or complete abolition of enzyme activity are associated with OCA1A, characterised by an almost complete absence of hair, skin and eye pigmentation (129, 130). Hypomorphic *TYR* variants in which mutant tyrosinase possess residual catalytic activity are associated with OCA1B, where affected individuals present with a milder phenotype with reduced levels of pigmentation (129, 130).

This chapter entails clinical and genomic investigations into the cause of OCA in four interlinking Amish families, which alongside functional studies and a review of previously published OCA genomic studies, highlights the pathogenicity of two common hypomorphic *TYR* variants p.(Ser192Tyr) and p.(Arg402Gln) that were previously largely considered non-pathogenic polymorphisms. Additionally, this chapter compiles genetic findings regarding the causes of OCA in 36 families from several communities in Pakistan. Together with a comprehensive literature review of all pathogenic gene variants associated with OCA in the Pakistani population, these findings permit a detailed understanding of the molecular spectrum of OCA in Pakistan.

Within this chapter, I was responsible for the interpretation and analysis of all collected clinical data for all affected individuals, including the development of clinical proformas to aid collaborating scientists acquire the best quality and most relevant clinical data for the Pakistan OCA families (families 5 - 40) described in chapter 3.3. I performed DNA extraction for a proportion of affected families (remaining DNA extraction largely completed by Joe Leslie, University of Exeter), primer design for amplification of *TYR*

exons 4 and 5 (primers for amplification of *TYR* exons 1-3 designed by Dr Gaurav Harlalka, University of Exeter) as well as for all variants identified in families 5 – 39, and completed all cosegregation studies. I was also responsible for analysing results of all exome sequencing and SNP mapping studies performed in this chapter. All literature reviews within this chapter were performed by myself, including the statistical analysis across different OCA cohorts in chapter 3.2, as well as the compilation of all pathogenic gene variants associated with OCA in the Pakistani population, detailed in appendix C.

3.2 Evidence that the Ser192Tyr/Arg402Gln in *cis* Tyrosinase gene haplotype is a disease-causing allele in oculocutaneous albinism type 1B (OCA1B)

3.2.1 Introduction

The apparent missing heritability in OCA is well described, with ~25-30% of clinically affected individuals lacking two clearly pathogenic sequence alterations within the same OCA gene; this proportion is higher in individuals with a partial OCA phenotype (129, 131). Several hypotheses have been proposed to explain this missing heritability, including variants in promoter or other regulatory elements, as well as epistatic or synergistic interactions between known genes (129, 132). Two *TYR* sequence variants [NM_000372.4:c.575C>A; **p.(Ser192Tyr)** or S192Y and c.1205G>A; **p.(Arg402Gln)** or R402Q], previously described as non-pathogenic polymorphisms due to their frequency in the general population (25% and 18% respectively), have been found to be enriched in cohorts of OCA patients with only one identified *TYR* mutation (126, 129, 133-141), leading to suggestions that these variants may in fact account for some of this missing heritability (117, 126, 127, 133, 134, 137, 142-145). This has however been disputed by others (136, 138, 146) who have hypothesised that these variants may be pathogenic only when present in *cis*, and inherited in biallelic fashion with a second deleterious *TYR* variant for tyrosinase activity to be sufficiently reduced to a level that will cause an OCA phenotype (132, 145, 147). This hypothesis is supported by increasing numbers of individuals with OCA reported to carry one pathogenic variant as well as the p.(Ser192Tyr) and p.(Arg402Gln) variants in *TYR* (132, 141, 148, 149). However, due to the high frequency of the p.(Ser192Tyr) and p.(Arg402Gln) variants in the general population, and the often small family sizes common to modern European populations, in many cases it has not been possible to obtain informative allele segregation to phase gene variants and prove inheritance of a *cis* p.(Ser192Tyr)/p.(Arg402Gln) haplotype in *trans* with the pathogenic *TYR* alteration in all affected individuals (145, 148). The remaining uncertainty in clinical interpretation of this haplotype limits its routine reporting in diagnostic gene panels. This has important diagnostic implications; designating the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype as pathogenic could substantially

increase the diagnostic yield by ~25-50% in albinism patient cohorts with missing heritability (132). This also further supports the hypothesis that the prevalence of OCA1, commonly quoted as ~1 in 40,000 (130), likely represents a substantial underestimation particularly amongst Caucasian populations with fair pigmentation (149).

This study entails extensive genetic studies stemming from the investigation of a large multigenerational extended Amish family, alongside functional studies, and a review of genotyped UK based albinism cohorts and of existing literature, to provide strong evidence supporting pathogenicity of the *TYR* p.(Ser192Tyr)/(Arg402Gln) in *cis* haplotype and its contribution to OCA1B in European populations.

3.2.2 Materials and methods

Amish patient ascertainment and molecular genetic analysis

Affected individuals and unaffected family members from four Ohio and Wisconsin Amish families with a common Ohio ancestry were recruited to this study with informed consent (Figure 3.1). A medical history was taken in all recruited family members, as well as detailed phenotyping of skin and hair pigmentation, particularly in the context of familial pigmentary background. A diagnosis of nystagmus was established in all affected individuals (apart from individual X:5), and further ophthalmic investigations including electroretinography and OCT were performed in selected individuals. A diagnosis of albinism was made following a set of diagnostic criteria proposed by Kruijt et al (119).

Blood/buccal samples were obtained with informed consent for DNA extraction (section 2.3.2). Exome sequencing (WES for individual IX:9 and Illumina TruSight™ One clinical exome sequencing panel for individual IX:22) was performed as described in section 2.3.5. Bioinformatic analysis of exome data was performed as per section 2.3.5, with additional filtering performed using the “Albinism or congenital nystagmus v1.0” PanelApp virtual gene panel (41 genes) (https://panelapp.genomicsengland.co.uk/panels/).

Primers were designed as described in section 2.3.3 to cover all coding exons and associated intron-exon junctions of *TYR* (Appendix Table D1). As the 3' region encompassing coding exons 4 and 5 of *TYR* shares high homology with a pseudogene (*TYRL*) (150), locus-specific amplification primers were designed for *TYR* exons 4 and 5 to prevent co-amplification of *TYR* and *TYRL* and subsequent misinterpretation of results. PCR and dideoxy sequencing was performed as described in section 2.3.3 to genotype and confirm appropriate segregation of candidate disease variants in all available affected and unaffected individuals.

Establishment of Tyr mutant cell lines, cell culture conditions and enzymatic activity assays

(Performed by Chelsea Norman and Aida Sanchez-Bretaño, University of Southampton)

The plasmid vector p3XFLAG-CMV-14 containing *TYR* cDNA was purchased from Addgene (Massachusetts, USA) and was initially deposited by Ruth Halaban (151). Upon arrival, sequencing revealed the p.(Ser192Tyr) (c.C575A) common population variant to be present. Site-directed mutagenesis was used to create the wild-type sequence (c.575C, p.192Ser) as well as the p.(Arg402Gln) variant. Site-directed mutagenesis was carried out using the non-strand displacing activity of Pfu DNA polymerase to incorporate and extend the mutagenic primers. The reaction mixture contained Phusion Pfu Polymerase and its buffer, forward and reverse primers (0.5 µM), dNTPs (200 µM), and the cDNA template. PCRs were performed in a total volume of 50 µl. Touch-down PCR conditions were set at 98°C for 30 sec followed by 30 cycles of 98°C for 10 sec, 45-72°C for 10 - 30 sec and 72°C for 15-30 sec, and a final extension step of 72°C for 5 - 10 min. The PCR product was treated with DpnI to digest the methylated parental DNA.

Purified mutated tyrosinase PCR products were employed to transform NEB® 5-alpha Competent E. coli (High Efficiency; New England Biolabs, UK) via heat shock method. Briefly, 50 µl of thawed cells were kept on ice and combined with approximately 100 ng of plasmid DNA and incubated for 30 min. The cell-DNA mixture was heat shocked at 42˚C for 30 sec and then placed on ice for 5 min. Cells were given S.O.C medium and incubated for an hour in a shaking incubator before being plated on ampicillin

selection (100 ug/ml) LB agar plates. After overnight incubation at 37˚C, single ampicillin resistant colonies were picked and grown in LB broth for approximately 16 hours, at which point the cells were pelleted by centrifugation and the DNA extracted. When the stocks were diminished, competent cells were produced through treatment with $CaCl_2$ and subsequently transformed using the heat shock method described above.

Human Embryonic Kidney 293 Freestyle (HEK293F) cells (Invitrogen, California, USA) were cultured in Freestyle culture medium (Invitrogen, California, USA) at 37˚C in a shaking incubator at 125 rpm with 8% CO_2. When cells reached a density of 1 $x10^6$ cells/ml, they were transfected with 30 µg of plasmids containing the p.(Arg402Gln) or p.(Ser192Tyr) mutations or co-transfected with both plasmids. The lipid-based reagent, 293fectin (60 µl) (ThermoFisher, UK), was diluted in Opti-MEM (ThermoFisher, UK) and incubated at room temperature for 5 min. DNA and 293fectin were combined, gently mixed and incubated at room temperature for 30 min before adding to cells. Then, cells were incubated in 6-well plates for 72 hours at 31°C or 37°C to reach 90% confluency, and the enzymatic activity assays were performed.

DOPA-oxidase activity was assessed in the different mutants. First, transfected cells from the different mutant clones were treated with L-DOPA, and the DOPA-oxidase activity was measured as the accumulation of the downstream product, dopachrome, following the manufacturer's protocol. Briefly, cells cultured in 6-well plates were lysed in NP40 Cell Lysis Buffer (ThermoFisher, UK) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) (in DMSO with a final concentration of 1%) and 1X protease and phosphatase inhibitor (Halt™ Phosphatase Inhibitor Cocktail, Thermo Fisher Scientific, UK), and protein concentration was measured by BCA assay (Thermo Scientific™ Pierce™ BCA Protein Assay Kit). Samples were then diluted into 4 µg/µl, and 50 µl or 30 µl sample aliquots were used for the DOPA assays. After adding the volume of the samples to 96-well plates, 150 µl of a phosphate buffer with L-DOPA 1 mM was added to the wells. Enzymatic activity was recorded as the absorbance of dopaquinone at 492 nm from the start of L-dopa treatment (0 min) and at 30 min intervals thereafter for a total of 180 min at both 31°C and 37°C. Assays were routinely performed in triplicate and the results are presented as the means of the independent assays +/- standard error.

Results of enzymatic activity at 180 min were normalized to wild type, with the values for wild type taken to be 100% of the expected enzymatic activity. One-way ANOVA was performed followed by a Sidak's post-hoc test. A probability level of at least $p<0.05$ was considered statistically significant (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$).

Evaluating the prevalence of TYR *p.(Ser192Tyr)/p.(Arg402Gln) haplotype in OCA and control cohorts*

A clinical cohort of affected individuals with nystagmus and/or albinism was retrospectively ascertained through the Southampton (161 individuals) and Salisbury (131 individuals) research databases. All individuals had been referred from a regional paediatric nystagmus clinic and recruited with informed consent. NGS (Illumina TruSight™ One clinical exome sequencing panel), alignment and filtering was performed as previously described in section 2.3.5. The genomic data were interrogated to ascertain frequency of the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype in this cohort. A literature review was also performed to evaluate the reported prevalence of the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype in additional published OCA cohorts. This was compared against an in-house exome database of 219 Amish individuals unaffected by OCA. Statistical analysis was performed using an established software package (R Core Team 2015; R Foundation for Statistical Computing, Vienna, Austria) (152).

3.2.3 Results

Clinical and genetic findings

A large multigenerational extended Amish family residing in Wisconsin (USA) with nine affected individuals all exhibiting nystagmus and variable levels of hair and skin hypopigmentation was investigated (Figure 3.1A; family 4). On the basis of a detailed medical history, assessment of skin and hair pigmentation, and ophthalmic investigations in selected affected individuals, a diagnosis of likely mild OCA was made in all affected individuals. Two additional Amish families with a total of four

affected individuals with a similar clinical phenotype were subsequently recruited (Figure 3.1A; families 2 and 3). In addition, a further Amish family with a single affected individual with OCA was recruited to the study (Figure 3.1A; family 4). This individual displayed clinical features consistent with a complete OCA phenotype, including pale skin and white/blonde hair and eyelashes, nystagmus, iris transillumination defects and foveal hypoplasia. Affected individuals were not noted to bruise or bleed easily, although specific haematological investigations were not performed. Clinical findings for all affected individuals are summarised in Table 3.1.

Table 3.1 Summary of clinical features observed in affected individuals in families 1 - 4 with OCA

Family (ID)	Nystagmus	Hair colour	Eye colour	Other ocular features	Other systemic features
1 (X:1)	✓	Blonde	Blue	Iris transillumination defects, depigmented fundus, foveal hypoplasia, alternating esotropia, optic disc hypoplasia	-
2 (X:2)	✓	NA	NA	Blunted foveal reflex, depigmented fundus	NA
3 (X:3)	✓	Blonde	Blue	Iris transillumination defects, blunted foveal reflex	-
3 (X:4)	✓	Dark blonde	Blue	transillumination defects, foveal hypoplasia, strabismus	-
3 (X:5)	NA	Strawberry blonde	NA	Blunted foveal reflex	-
4 (IX:9)	✓	Blonde	Blue	Pale fundi, iris transillumination defects, foveal hypoplasia, myopia, strabismus. Nyctalopia, photosensitivity and peripheral VF loss with normal ERG	
4 (IX:10)	✓	Pigmented	Blue	-	Mild learning difficulties
4 (IX:12)	✓	Light brown	NA	-	-
4 (IX:14)	✓	Dark brown	Blue	Pale fundi	-
4 (IX:15)	✓	Pigmented	Blue	-	-
4 (IX:16)	✓	NA	NA	NA	NA
4 (IX:20)	✓	Blonde	Blue	-	-
4 (IX:22)	✓	White/ blonde	Blue	-	-
4 (X:15)	✓	Brown	Brown	Myopia	Neonatal intraventricular haemorrhage

Abbreviations: ERG, electroretinogram; NA, information not available; VF, visual field. The (✓) and (-) symbols indicate the presence of absence of a feature in an affected subject respectively

All affected individuals were diagnosed with hypomorphic albinism on the basis of a positive molecular diagnosis as well as the presence of either one major criterion or two minor criteria described by Kruijt et al *(119). Major criteria includes foveal hypoplasia ≥ grade 2, crossed asymmetry on visual evoked potential testing and ocular hypopigmentation (either iris transillumination or fundus hypopigmentation ≥ grade 2), whilst minor criteria includes nystagmus, hypopigmentation of skin and hair, and grade 1 fundus hypopigmentation and/or foveal hypoplasia.*

Exome sequencing was initially performed in two affected individuals in family 4 (individuals IX:9 and IX:22) for targeted evaluation using the "Albinism or congenital nystagmus v1.0" PanelApp virtual gene panel (41 genes). Subsequently, variants predicted to have a functional consequence (including CNVs) located genome-wide were identified and filtered according to allele frequency (gnomAD MAF of <0.01). This identified only a single plausible candidate disease variant in both individuals, a heterozygous *TYR* missense variant (GRCh38) chr11:g.89178708T>G; NM_000372.4:c.755T>G; p.(Met252Arg) or M252R. The p.Met252 amino acid residue is located in the catalytic domain of the tyrosinase protein, and is conserved across a variety of vertebrate species (Figure 3.1B). This variant was absent in gnomAD and Genome Project population databases, although it was present in an Amish control exome dataset (allele frequency 0.0023) in heterozygous form only. *In silico* analysis of the p.(Met252Arg) variant using SIFT, PolyPhen-2 and PROVEAN predicted the variant to be deleterious, possibly damaging and deleterious. This variant has been reported in compound heterozygous form [with a previously reported p.(Arg217Trp) variant] in a single individual with OCA (141), and is considered to be likely pathogenic. Exome sequencing did not identify any additional candidate single nucleotide or structural disease variants in any OCA-associated genes.

To explore this apparent missing heritability, targeted dideoxy sequencing of all coding regions and intron-exon junctions of the *TYR* gene was performed in these two individuals. This confirmed the presence of the p.(Met252Arg) variant, and also identified a further two *TYR* missense variants (GRCh38) chr11:g.89178528C>A; NM_000372.4:c.575C>A; p.(Ser192Tyr) (S192Y) and (GRCh38) chr11:g.89284793G>A; NM_000372.4:c.1205G>A; p.(Arg402Gln) (R402Q) in the same two individuals, excluded from the exome sequencing analysis based on population allele frequencies of 0.25 and 0.18 respectively. Segregation of all three *TYR* variants in all Amish families (families 1-4) is shown in Figure 3.1A, which

demonstrates that the p.(Ser192Tyr)/p.(Arg402Gln) variants are linked in *cis* and inherited in a compound heterozygous fashion with p.(Met252Arg) [which itself occurs in cis with p.(Arg402Gln)] in all affected individuals except for a single affected individual with OCA (family 1; individual X:1), found to be homozygous for p.(Met252Arg) through targeted dideoxy sequencing. Individuals compound heterozygous for *TYR* p.(Met252Arg) and p.(Ser192Tyr)/p.(Arg402Gln) alleles displayed clinical features suggestive of partial albinism with variable skin and hair depigmentation, whilst the individual homozygous for the *TYR* p.(Met252Arg) variant displayed features of classical OCA including nystagmus, iris transillumination defects, a depigmented fundus and foveal hypoplasia (Table 3.1). Notably, individuals carrying the *TYR* p.(Met252Arg) variant on one allele and only the p.(Arg402Gln) or the p.(Ser192Tyr) variant on the other allele were apparently unaffected with no clinical features of OCA (individuals VIII:9, IX:2, IX:21, X:6, X:8, IX:1 and IX:4, Figure 3.1A).

Additive temperature sensitive effects of p.(Ser192Tyr) (S192Y) and p.(Arg402Gln) (R402Q) variants on TYR enzymatic activity

The *TYR* p.(Arg402Gln) variant alone has previously been proposed to contribute to OCA when inherited in *trans* with a pathogenic *TYR* variant (127, 133-135, 141-144, 153). Pedigree analysis however appears to dispute this, with five individuals compound heterozygous for the pathogenic *TYR* p.(Met252Arg) variant as well as the p.(Arg402Gln) variant and yet showing no clinical features of OCA (individuals VIII:9, IX:2, IX:21, X:6 and X:8; Figure 3.1A). At the same time, 13 individuals who were compound heterozygous for *TYR* p.(Met252Arg) and p.(Ser192Tyr)/p.(Arg402Gln) alleles all displayed clinical features of partial albinism, suggesting an additive impact of the p.(Ser192Tyr) and p.(Arg402Gln) variants on tyrosinase function. To investigate this further, functional experiments were designed to study and quantify the effects of the p.(Ser192Tyr) and p.(Arg402Gln) variants both independently and in combination, compared to wild-type tyrosinase enzyme.

Figure 3.1C shows the DOPA-oxidase activity for all tyrosinase mutants analysed from 0 mins to 180 mins at 31°C and 37°C. At 37°C, a slight decrease in DOPA-oxidase activity of the p.(Ser192Tyr) mutants, and an almost total loss of DOPA-oxidase

activity in the p.(Arg402Gln) mutants and p.(Ser192Tyr)/p.(Arg402Gln) double mutants was observed. Tyrosinase activity recorded at 31°C was more variable across the different mutant groups, with significant decrease in activity in all *TYR*-mutant cell lines compared to wild type. For all the *TYR*-mutant cell lines, the p.(Ser192Tyr)/p.(Arg402Gln) double mutants showed the most reduced tyrosinase activity, followed by p.(Arg402Gln) mutant, with the p.(Ser192Tyr) mutant least affected. There was a statistically significant difference between all three mutant groups, indicative of a cumulative effect of both p.(Ser192Tyr) and p.(Arg402Gln) mutations on tyrosinase activity.

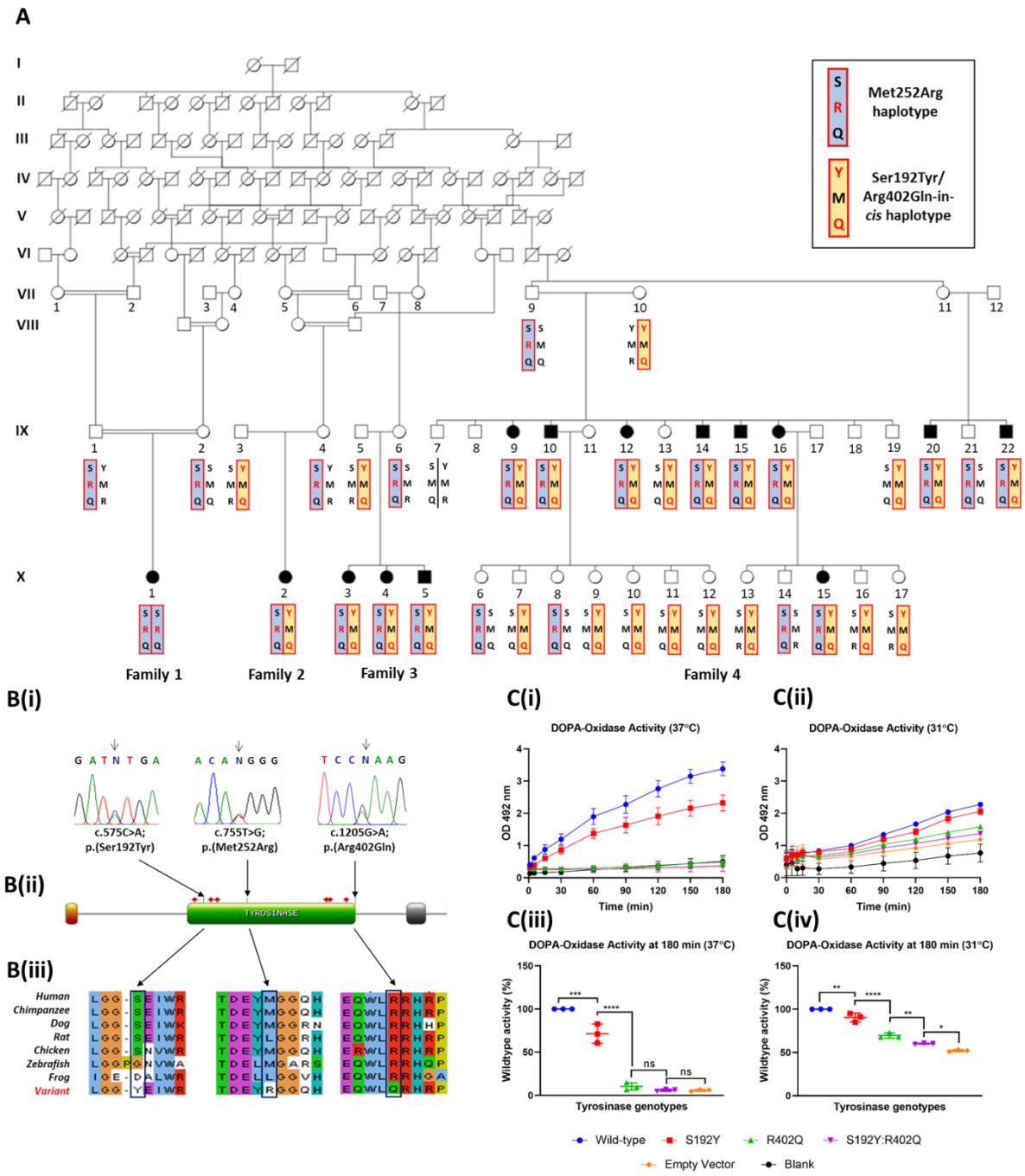


Figure 3.1 Pedigree diagrams, *TYR* genotype and functional data

(A) Pedigree diagram of families 1 - 4 showing segregation of the TYR *p.(Ser192Tyr), p.(Met252Arg), and p.(Arg402Gln) variants. The two disease-causing haplotypes are shaded; the p.(Met252Arg) haplotype in blue, and the p.(Ser192Tyr)/p.(Arg402Gln) in* cis *haplotype in yellow, with the relevant amino acid change highlighted in red. All affected individuals inherited both the p.(Met252Arg) and p.(Ser192Tyr)/p.(Arg402Gln) in* cis *haplotypes in compound heterozygous fashion, apart from individual X:1, who displayed a more severe albinism phenotype and was found to be homozygous for the p.(Met252Arg) variant. No unaffected individuals inherited both disease-associated haplotypes, confirming segregation with disease in the family.*
(B)(i) Sequence chromatograms showing the TYR *c.575C>A; p.(Ser192Tyr), c.755T>G; p.(Met252Arg) and c.1205G>A; p.(Arg402Gln) variants in heterozygous form (ii) Schematic localisation of* TYR *p.(Ser192Tyr), p.(Met252Arg) and p.(Arg402Gln) variants within the*

catalytic tyrosinase domain of the TYR polypeptide. The p.(Ser192Tyr) and p.(Arg402Gln) variants are located at or near the copper-containing catalytic binding sites (the red diamonds denote the histidine residues that bind to copper atoms and hence structurally coordinate the positions of the metal binding sites) (iii) Conservation of the TYR *p.(Ser192Tyr), p.(Met252Arg) and p.(Arg402Gln) variants across species.*
(C) Tyrosinase activity in wild-type, p.Ser192Tyr (S192Y) mutant, p.Arg402Gln (R402Q) mutant and double mutant HEK293 cells. Absorbance of dopaquinone, a product synthesised by the transformation of L-DOPA by tyrosinase was quantified as a measure of tyrosinase activity in wild-type and TYR-mutant cell lines. Cumulative production of dopaquinone (i, ii)) was quantified from the start of the L-DOPA treatment (0 min) to 180 min. Statistical differences between cell lines were analysed at 180 min (iii, iv). Data is shown as mean ± SEM and statistically significant differences between groups are indicated by asterisks (p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001); ns = not significant*

Enrichment of the TYR *p.(Ser192Tyr)/p.(Arg402Gln) haplotype in OCA and control cohorts*

Interrogation of a clinical cohort of 161 affected individuals with nystagmus and/or albinism (Southampton cohort) [including individuals previously reported by Norman *et al* and O'Gorman *et al* (132, 145)] identified 71 individuals with two pathogenic or likely pathogenic variants (molecularly diagnosed including *TYR*, *OCA2*, *GPR143* and *PAX6* genes), 51 individuals carrying only a single likely disease-associated *TYR* variant with no candidate pathogenic variants identified in other OCA genes (missing heritability), and 39 individuals with no disease-associated *TYR* variants. All patients were sequenced using either using the "Albinism or congenital nystagmus v1.0" PanelApp gene panel (41 genes) (https://panelapp.genomicsengland.co.uk/panels/) or a broader research panel as previously described (132, 145). CNV analysis was not performed. Of these, 2 of the 71 individuals in the molecularly diagnosed group and 49 of the 51 individuals in the missing heritability group were found to have a genotype consistent with the presence of the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype (i.e. individuals who were homozygous or heterozygous for both these variants) (Table 3.2); this information was unavailable for the 39 molecularly undiagnosed individuals in this clinical cohort.

A review of seven published OCA cohorts with missing heritability (i.e. individuals in whom only a single pathogenic *TYR* variant has been identified), together with our study cohort, found that approximately half of all affected individuals (50.7%) had a genotype consistent with the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype (Table 3.2). This is markedly enriched compared to molecularly diagnosed *TYR* OCA cohorts

(2.0%), as well a control cohort of 219 Amish individuals with no OCA diagnoses (31/219 or 16.9%; Pearson's Chi-squared test, $p < 2.2 \times 10^{-16}$). These findings strongly suggest that the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype contributes to the OCA phenotype.

Additionally, 49 affected individuals in the Southampton and Salisbury missing heritability study cohorts identified as carrying only a single pathogenic or likely pathogenic *TYR* variant were also found to harbour homozygous or heterozygous *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants; of these, familial segregation was performed in 41 individuals and their parents to assess the phase. In 23 individuals, this confirmed that the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants were inherited in *cis*, and this haplotype was in *trans* to the previously identified pathogenic or likely pathogenic *TYR* variant (Table 3.3). For the remaining 18 cases, definitive segregation was not possible. Notably, no case was identified in which segregation showed that p.(Ser192Tyr) and p.(Arg402Gln) were not in *trans* with the pathogenic or likely pathogenic variant.

In five of the seven published OCA cohorts with missing heritability reviewed, it was possible to determine the *cis/trans* phase of the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants in a proportion of individuals reported (136, 139, 141, 148, 149) (Table 3.3); in the remaining individuals this was not possible due to familial samples being unavailable for segregation analysis, or uninformative segregation results (owing to the high allele frequency of the p.(Ser192Tyr) and p.(Arg402Gln) *TYR* variants in the general population). For the remaining two studies of OCA cohorts with missing heritability, the *cis/trans* phase of the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants could not be determined from the reported genotypes (126, 134). There were 41 OCA individuals with missing heritability from these five studies in whom the p.(Ser192Tyr)/p.(Arg402Gln) haplotype was possible, and where the *cis/trans* phase of the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants could also be determined. In accordance with the findings from our local research cohorts, together with this additional informative cohort derived from five published studies, the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype segregated in *trans* with the pathogenic *TYR* variant in all cases (amounting to 25.5% of total missing heritability cases; Table 3.3). Taken together with the findings in Table 3.2, this suggests that the

p.(Ser192Tyr)/p.(Arg402Gln) haplotype completes the molecular diagnosis in ~25-50% of OCA individuals with missing heritability.

Table 3.2 Prevalence of *TYR* p.(Ser192Tyr)/S192Y and p.(Arg402Gln)/R402Q variants in OCA cohorts

	OCA cohorts with missing heritability (individuals with only 1 TYR pathogenic or likely pathogenic variant identified)								Molecularly diagnosed OCA1 cohorts (individuals with 2 TYR pathogenic or likely pathogenic variant identified)			
	This study*	Hutton & Spritz 2008a	Hutton & Spritz 2008b	Oetting 2009	Ghodsinejad Kalahroudi 2014	Lasseaux 2018	Gronskov 2019	Campbell 2019	Hutton & Spritz 2008b	Oetting 2009	Ghodsinejad Kalahroudi 2014	Gronskov 2019
Phenotype	Nystagmus and/or albinism	AROA/ mild OCA	OCA	OCA1	OCA1	Nystagmus and/or absence of fovea	Albinism (OCA, AROA or OA)	Nystagmus and at least one other ocular feature of albinism, no skin hypopigmentation	OCA	OCA1	OCA1	Albinism (OCA, AROA or OA)
Country (ethnicity)	England	(Caucasian)	USA, Canada, Northern Europe (non-Hispanic/ Latino Caucasians)	NA	(Iranian)	France	Scandinavia (Scandinavian)	England	USA, Canada, Northern Europe (non-Hispanic/ Latino Caucasians)	NA	(Iranian)	Scandinavia (Scandinavian)
No of individuals in cohort	51	20	13	3	6	158	29	4	71	9	19	2
No of individuals hom or het for both *TYR* S192Y & R402Q	49	1	3	2	0	64	21	4	2	0	0	0
Proportion of study cohort where S192Y/R402Q haplotype is possible	49/51 (96.1%)	1/20 (5.0%)	3/13 (23.1%)	2/3 (66.7%)	0/6 (0%)	64/158 (40.5%)	21/29 (72.4%)	4/4 (100%)	2/71 (2.8%)	0/9 (0%)	0/19 (0%)	0/2 (0%)
Combined proportion where S192Y/R402Q haplotype is possible	144/284 (50.7%)								2/101 (2.0%)			

**This cohort includes individuals previously reported in Norman et al and O'Gorman et al*

Abbreviations: AROA, autosomal recessive ocular albinism; het, heterozygous; hom, homozygous; OCA, oculocutaneous albinism; OA, ocular albinism; no, number

Table 3.3 Potential contribution of *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype to molecular diagnoses in OCA cohorts with missing heritability

This includes individuals with only 1 *TYR* pathogenic or likely pathogenic variant identified

Study	This study*	Oetting 2009	Ghodsinejad Kalahroudi 2014	Lasseaux 2018	Gronskov 2019	Campbell 2019
Phenotype	Nystagmus and/or albinism	OCA1	OCA1	Nystagmus and/or absence of fovea	Albinism (OCA, AROA or OA)	Nystagmus and at least one other ocular feature of albinism, no skin hypopigmentation
Country (ethnicity)	England	NA	(Iranian)	France	Scandinavia (Scandinavian)	England
Number of individuals in cohort	51	3	6	158	29	4
Number of individuals hom or het for both *TYR* S192Y and R402Q, where S192Y/R402Q haplotype is possible	49	2	0	64	21	4
Number of individuals in whom it was possible to determine the phase of *TYR* S192Y, R402Q and pathogenic or likely pathogenic variants ("informative cohort")	23	2	6	31	6	2
Number of individuals in whom *TYR* S192Y and R402Q were in *cis*, and in *trans* to pathogenic or likely pathogenic *TYR* variant in the informative cohort	23	2	0	31	6	2
Proportion of "informative cohort" where S192Y/R402Q haplotype is possible and molecular diagnoses due to *TYR* S192Y/R402Q haplotype in *trans* to pathogenic or likely pathogenic *TYR* variant	23/23 (100%)	2/2 (100%)	S192Y/R402Q haplotype not possible in any individuals in study	31/31 (100%)	6/6 (100%)	2/2 (100%)
Combined proportion of "informative cohort" where molecular diagnoses is due to *TYR* S192Y/R402Q haplotype in *trans* to pathogenic or likely pathogenic *TYR* variant				64/64 (100%)		
Proportion of total cohort where molecular diagnoses due to *TYR* S192Y/R402Q haplotype in *trans* to pathogenic or likely pathogenic *TYR* variant	23/51 (45.1%)	2/3 (66.7%)	0/6 (0%)	31/158 (19.6%)	6/29 (20.7%)	2/4 (50%)
Combined proportion of total cohort where molecular diagnoses is due to *TYR* S192Y/R402Q haplotype in *trans* to pathogenic or likely pathogenic *TYR* variant				64/251 (25.5%)		

**This cohort includes individuals previously reported in Norman et al and O'Gorman et al*
Abbreviations: AROA, autosomal recessive ocular albinism; het, heterozygous; hom, homozygous; OCA, oculocutaneous albinism; OA, ocular albinism.

3.2.4 Discussion

The pathogenicity of *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants and their contribution to the OCA phenotype, either in isolation or when linked in *cis*, has been heavily debated in many studies (117, 126, 127, 133, 134, 136-138, 142-146). As such, these *TYR* variants are variably reported by clinical testing laboratories and potentially excluded, even when shown to be in *cis*. Here genomic and functional data, initiated by a search for the cause of OCA in a number of Amish families, provide irrefutably strong evidence that the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants are pathogenic when in *cis*. The increased frequency of the *TYR* p.(Met252Arg) variant in the Amish community, likely due to founder effects and endogamy together with the large family sizes typical within the community, permitted empowered co-segregation studies able to determine the haplotype, phasing and inheritance of the common p.(Ser192Tyr) and p.(Arg402Gln) *TYR* variants together with the p.(Met252Arg) variant in a large number of related individuals for the first time.

A number of previous studies have shown that the *TYR* p.(Arg402Gln) variant is present at increased frequency in OCA patients in whom only a single clearly deleterious sequence alteration in *TYR* has been detected (126, 129, 133-141, 149), and is proposed to contribute to OCA when inherited in *trans* with a pathogenic *TYR* variant (127, 133-135, 141-144, 153). This has been disputed by Oetting *et al*, who describe 10 unaffected parents of OCA individuals who carried a combination of both a *TYR* pathogenic variant as well as the p.(Arg402Gln) variant in *trans*, and yet showed no clinical features of OCA (136). Recently, however, molecular and phenotypic analysis of a large cohort of 268 French OCA1 individuals suggests that compound heterozygosity for a known pathogenic *TYR* variant and the hypomorphic p.(Arg402Gln) variant may be associated with a variable but generally mild form of albinism that may remain clinically undiagnosed, with the authors supporting consideration of p.(Arg402Gln) as a mildly pathogenic, but definitely pathogenic, *TYR* variant when associated in trans with another pathogenic variant (154). It should be noted that the *TYR* p.(Arg402Gln) variant is not generally thought to be sufficiently deleterious to cause OCA in homozygous form, with a homozygous

prevalence of 7.47% in population databases (non-Finnish European population, gnomAD v2.1.1), as well as several studies that have reported on individuals homozygous for the p.(Arg402Gln) with no clinical features of albinism (136, 138, 155).

It has been proposed that a second *TYR* variant, p.(Ser192Tyr), acting in *cis* with the p.(Arg402Gln) variant, may have an additive effect producing a greater reduction in enzyme activity compared to each variant individually (145, 147), although this too has been disputed (154). Our study demonstrates that inheritance of either variant individually in compound heterozygous form with the deleterious p.(Met252Arg) variant is insufficient to result in a clinically significant OCA phenotype (individuals VIII:9, IX:1, IX:2, IX:4, X:6 and X:8; Figure 3.1A), although we acknowledge that subtle asymptomatic clinical findings such as mild iris transillumination or foveal hypoplasia were not specifically excluded. The p.(Ser192Tyr) and p.(Arg402Gln) variants are believed to have arisen independently on different ancestral haplotypes (156), and although these variants individually in Caucasian populations are common with allele frequencies of 36% and 27% respectively (gnomAD v2.1.1), their combined presence in *cis* on a recombinant haplotype is relatively rare, predicted to be between 1.1% to 1.9% in European populations (145, 147, 149). This study, alongside other previous studies (126, 132, 136, 141, 145, 148, 149), provides strong support to show that the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype is enriched in Caucasian OCA cohorts with missing heritability (Table 3.2), and contributes to an OCA1B diagnosis when inherited in *trans* with a second deleterious *TYR* variant (Table 3.3), particularly in individuals with lower pigmentary backgrounds, who may be more susceptible to the damaging effects of hypomorphic variants (135, 157).

Given the number of apparently unaffected individuals homozygous for the p.(Ser192Tyr)/p.(Arg402Gln) haplotype reported in the literature (147-149), the penetrance of the p.(Ser192Tyr)/p.(Arg402Gln) haplotype might appear to be incomplete, confounding the argument that it is a pathogenic allele. The phenotype of OCA2 can be modified by variants in *MC1R* (158) or *TYRP1* (159), and the phenotype of OCA3 by haploinsufficiency of *OCA2* (160);

additionally, possible digenic inheritance in OCA involving combinations of *TYR*, *OCA2* and *SLC45A2* variants has been proposed (161, 162). Similarly, the apparent reduced penetrance of the p.(Ser192Tyr)/p.(Arg402Gln) haplotype may relate to the modifying effects of sequence variants in genes encoding other melanosomal proteins, although other genetic and molecular studies would be required to confirm this. However, we propose that individuals homozygous for the hypomorphic p.(Ser192Tyr)/p.(Arg402Gln) allele have instead a consistent but mild phenotype which is easily missed by incomplete phenotyping. In support of this, our studies identified five individuals with a clinical diagnosis of 'possible hypomorphic' OCA who were homozygous for *TYR* p.(Ser192Tyr)/p.(Arg402Gln), with no other known or likely *TYR* or other OCA gene-associated variants identified (Table 3.4). All were noted to have foveal hypoplasia on OCT investigation, but most had very mild, if any, other OCA features. Additionally, an apparently unaffected relative in our study was also identified as homozygous for *TYR* p.(Ser192Tyr)/p.(Arg402Gln). Despite the absence of nystagmus or any other pigmentary phenotype in this unaffected individual and visual acuities of 0.1 and 0.08 LogMAR (right and left eye respectively), further detailed clinical investigation identified very mild iris transillumination and significant foveal hypoplasia (Table 3.4 and Figure 3.2). A review of all affected and apparently unaffected individuals homozygous for both *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants in the literature and our study cohorts identified 13 affected individuals in three studies, and foveal hypoplasia as well as iris transillumination was documented in all these individuals (148, 149) (Table 3.4). It therefore seems likely that individuals homozygous for both common variants and thus the hypomorphic p.(Ser192Tyr)/p.(Arg402Gln) allele, have such a mild phenotype that they can easily go unidentified and unreported due to minimal effects on visual function or clear features of albinism.

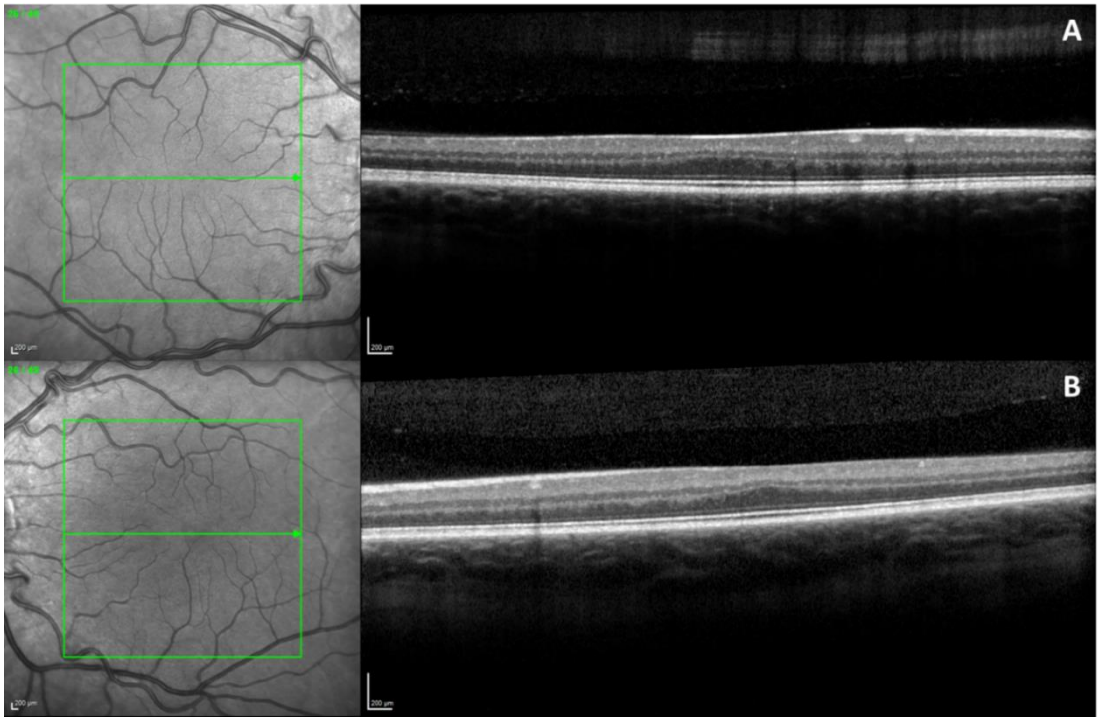


Figure 3.2 Foveal hypoplasia in individual homozygous for *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype

SD-OCT (Spectral domain optical coherence tomography; Spectralis-OCT, Heidelberg Engineering, Heidelberg, Germany) image of right (A) and left (B) eyes showing grade 3 foveal hypoplasia in an apparently unaffected individual homozygous for both TYR *p.(Ser192Tyr) and p.(Arg402Gln) variants [foveal hypoplasia graded following structural grading system based on OCT data proposed by Thomas* et al *(163)]*

The *TYR* p.(Arg402Gln) variant is located near the copper-containing catalytic binding site CuB, and functional studies have shown that this amino acid alteration results in an enzyme with decreased thermal stability, disrupted copper binding and reduced catalytic activity, thought to be mediated by decreased protein stability resulting in increased retention of the mutant tyrosinase protein as an unprocessed and misfolded glycoform in the endoplasmic reticulum (ER) (147, 155, 164-169). The *TYR* p.(Ser192Tyr) variant is located within the copper-containing catalytic binding site CuA, and has been shown to reduce tyrosinase enzymatic activity and melanocyte pigment production independent of the p.(Arg402Gln) variant (147, 170, 171). Genome-wide association studies have identified associations with skin, hair and eye pigmentation for both p.(Ser192Tyr) and p.(Arg402Gln) variants (172-176), suggesting these variants have a role in normal pigmentary variation, and that the double-variant p.(Ser192Tyr)/p.(Arg402Gln) haplotype appears to

show an additive effect on these pigmentary phenotypes compared to each variant individually (147). It is difficult however from literature review alone to quantify the functional effects of the p.(Ser192Tyr) and p.(Arg402Gln) variants, both independently and in combination, compared to wild-type tyrosinase enzyme. This issue arises from the historical use of the human *TYR* expression construct pcTYR containing the p.(Ser192Tyr) variant to study the effects of "wild-type" tyrosinase activity (155, 177). Computational approaches to TYR functional activity, based on protein flexibility and dynamic properties, suggest that the p.(Ser192Tyr) and p.(Arg402Gln) variants both result in a TYR protein that is less stable and has reduced enzyme activity compared to a wild-type molecule; the combined effect of having both changes together in a single TYR molecule however has not been previously investigated (178). This study now shows for the first time a thermosensitive additive decrease in enzymatic function of the double-variant p.(Ser192Tyr)/p.(Arg402Gln) TYR protein compared to each variant acting individually (Figure 3.1C), lending further support to pathogenicity of the p.(Ser192Tyr)/p.(Arg402Gln) haplotype. Homology modelling of tyrosinase protein structure does not appear to show a direct interaction between the 192 and 402 amino acid residues (149), and therefore this additional reduction in enzyme function in the double-mutant protein may instead be mediated by a combination of increased ER retention of the misfolded mutant protein [caused by p.(Arg402Gln) reducing protein stability] and reduced enzyme activity of any released mutant protein [possibly resulting from steric hindrance effects of p.(Ser192Tyr) affecting the CuA binding site] (170), as proposed by Gronskov *et al* (149).

Subcellular localisation studies have determined that disease-associated *TYR* variants commonly result in near absolute and irreversible ER retention of the mutant protein. The p.(Arg402Gln) variant however results in a thermosensitive tyrosinase protein that is retained in the ER at higher temperatures, but is able to partially exit the ER at lower, more permissive temperatures (164, 167, 169). Homozygosity for the p.(Ser192Tyr)/p.(Arg402Gln) haplotype may therefore still permit sufficient quantities of mutant tyrosinase to reach the inner surface of the melanosomal membrane, where the mutant protein is still able to participate in protein-protein interactions with other melanosomal proteins

involved in melanogenesis, such as TYRP1 and TYRP2 (179, 180), resulting in a less severe functional impact and a milder pigmentary phenotype that may not always be clinically significant. This thermosensitivity of the double variant mutant TYR protein also provides a compelling explanation for our novel discovery of a consistent foveal hypoplasia phenotype in individuals who are homozygous for both p.(Ser192Tyr)/p.(Arg402Gln) *TYR* variants, as higher temperatures within the developing eye may result in a larger impact of these variants on tyrosinase function (133), whilst lower temperatures at the skin and extremities instead results in greater preservation of mutant protein function and a milder and more variable pigmentary phenotype.

Together, our studies define the genotype, biochemical and phenotype correlation of the p.(Met252Arg) and p.(Ser192Tyr)/p.(Arg402Gln) *TYR* variants and collectively demonstrate that the in *cis* p.(Ser192Tyr)/p.(Arg402Gln) allele is pathogenic. As such, the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) haplotype should be included as a pathogenic allele in future and retrospective genetic diagnoses of OCA, supporting the idea for a review of all previously undiagnosed OCA cases where these variants have been excluded. Reporting of the p.(Ser192Tyr)/p.(Arg402Gln) genotype in individuals in whom only a single deleterious *TYR* variant has been identified could permit a 25-50% uplift in confirmatory molecular diagnoses (when phase has been determined) in this diagnostically challenging patient group (Tables 3.2 and 3.3). Additionally, for patients with an albinism phenotype but no apparent variants in albinism genes, consideration of these variants when identified in *cis* as a pathogenic allele in its own right may also help provide clinical direction. For example, in individuals heterozygous for this allele, alternative diagnoses such as syndromic albinism might be considered less likely, as they would be considered ‘at least a carrier of a pathogenic OCA1B allele’, and genomic data may be re-examined in a targeted fashion to search for further non-coding splice or structural variants in the *TYR* gene. Additionally, in individuals with a very mild albinism phenotype or isolated foveal hypoplasia, identification of this pathogenic allele in homozygous form may provide the molecular diagnosis, ending their diagnostic odyssey.

Achieving an accurate molecular diagnosis will bring about important benefits in affected individuals and their families. It allows accurate prognostic information and family counselling to be provided, avoids the need for further invasive investigations to confirm the clinical diagnosis or rule out syndromic forms of the disease or masquerading conditions, and has important therapeutic implications, given the emerging therapies currently under development and in clinical trials for OCA (181, 182).

Table 3.4 Review of individuals homozygous for both *TYR* p.(Ser192Tyr) and p.(Arg402Gln)

Individuals with an alternative molecular diagnosis responsible for the albinism phenotype were excluded from this review

Study	Cohort	No of individuals homozygous for both *TYR* p.(Ser192Tyr) and p.(Arg402Gln)	Comments
Individuals with albinism with no alternative molecular diagnosis and homozygous for p.(Ser192Tyr)/p.(Arg402Gln)			
Gronskov 2019 (149)	93 individuals with a clinical diagnosis of albinism; diagnostic criteria included nystagmus, reduced visual acuity, iris translucency, fundus hypopigmentation, and foveal hypoplasia	6 individuals • 5 with clinical diagnosis of AROA • 1 with clinical diagnosis of OCA	• All 6 individuals were first investigated by sequential Sanger sequencing of 6 genes only (*TYR*, *OCA2*, *TYRP1*, *SLC45A2*, *LRMDA*, *GPR143*) • 3 of the 6 individuals were further investigated by whole genome sequencing and data analysis of *TYR* genomic region
Campbell 2019 (148)	12 individuals with nystagmus, at least one other ocular feature of albinism, and no apparent skin hypopigmentation in the context of their family	2 individuals (Proband 7 and 8; not related)	Proband 7 • Genomic data analysis limited to nystagmus and foveal hypoplasia gene panel (26 genes evaluated) • Clinical information: iris transillumination, foveal hypoplasia, fundal hypopigmentation, probable VEP crossed asymmetry
			Proband 8 • Genomic data analysis limited to ocular/oculocutaneous albinism gene panel (18 genes evaluated) • Rare, likely pathogenic *TYRP1* variant [c.208 G>A, p.(Ala70Thr)] also identified in this individual • Affected brother with nystagmus, who was homozygous for both *TYR* p.(Ser192Tyr) and p.(Arg402Gln) but WT for the *TYRP1* variant • Clinical information: iris transillumination, foveal hypoplasia, fundal hypopigmentation, VEP not performed
This study	Salisbury cohort: 130 individuals with a clinical diagnosis of nystagmus and/or albinism All were investigated with Illumina TruSight™ One	5 individuals	W1703356 • Foveal hypoplasia noted on OCT • No other *TYR* or albinism-associated variant identified
			W1919237 • Afoveate adult with cystic fibrosis and nyctalopia

	clinical exome sequencing followed by additional filtering using virtual gene panel analysis using the "Albinism or congenital nystagmus v1.0" PanelApp gene panel		• No other *TYR* or albinism-associated variant identified • Biallelic *ABCA4* variants also identified and further investigation with ERG planned; however foveal hypoplasia is not a known feature of *ABCA4*-associated retinal dystrophy
			W2002293 • Possibly afoveate (poor quality OCT scans) • No other *TYR* variant or albinism-associated variant identified • Maternal uncle apparently affected with nystagmus
			W1905299 • Afoveate, pale skin and hair, good albinism phenotype • No other *TYR* variant or albinism-associated variant identified
			W1817121 • Afoveate; sister also afoveate but with no nystagmus • No other *TYR* variant or albinism-associated variant identified
Unaffected individuals homozygous for p.(Ser192Tyr)/p.(Arg402Gln)			
Jagirdar 2014 (147)	2 genetic epidemiological longitudinal studies: Brisbane Twin Nevus Study (1155 nuclear families) and Queensland Familial Melanoma Project (1211 melanoma cases)	2 individuals	Both individuals fair skinned with fair/blond hair and blue eye colour, no clinical diagnosis of OCA (but ocular examination not performed)
Campbell 2019 (148)	100,000 genomes pilot dataset of 4046 individuals with no clinical features suggestive of albinism	2 individuals	No detailed ocular phenotyping available
This study	Southampton cohort: 161 probands with nystagmus and/or albinism and relatives	1 individual	This apparently unaffected individual was father to two affected siblings who were both clinically diagnosed with nystagmus and/or OCA (see below). Both affected siblings were heterozygous for a known pathogenic *TYR* variant, and also homozygous for both *TYR* p.(Ser192Tyr) and p.(Arg402Gln). This unaffected parent was asymptomatic but was noted to have with mild iris transillumination and foveal hypoplasia on OCT (see Figure 3.2) in the absence of nystagmus or a pigmentary phenotype, and had normal visual acuities of 0.1 and 0.08 LogMAR (right and left eye respectively).
This study	Amish Exome Database: Control exome database of 219 Amish individuals unaffected by OCA	2 individuals	No detailed ocular phenotyping available

Individuals with albinism likely due to heterozygous deleterious *TYR* variant in combination with homozygosity for p.(Ser192Tyr)/p.(Arg402Gln)			
This study	Southampton cohort: 161 probands with nystagmus and/or albinism and relatives	2 individuals	Affected siblings, both also heterozygous for a known pathogenic *TYR* variant
This study	Salisbury cohort: 130 individuals with a clinical diagnosis of nystagmus and/or albinism	1 individual	W1809902 Also heterozygous for a *TYR* pathogenic variant
Lasseaux 2018 (141)	990 index patients with at least one of the main characteristic ocular features of albinism - either nystagmus or an absence of the fovea	1 individual	TYR/R402Q-P13 Also heterozygous for a known pathogenic *TYR* c.1118C>A; p.(Thr373Lys) variant (166)

Abbreviations: AROA, autosomal recessive ocular albinism; ERG, electroretinogram; OCA, oculocutaneous albinism; OCT, optical coherence tomography; VEP, visual evoked potential; WT, wild type.

3.3 Genetic spectrum of OCA in Pakistan

3.3.1 Introduction

In Pakistan, geographical constraints and marriage patterns within communities may give rise to genetic isolates in which an increased frequency of certain disease-associated founder mutations may occur (183). Knowledge of the specific spectrum and frequency of such genetic variants within different regions is fundamental to the development of effective and more tailored diagnostic genetic testing strategies, targeting variants relevant to a particular population. This will permit rapid and cost-efficient screening and diagnostic assays that will allow accurate disease diagnosis, improved carrier detection and appropriate counselling for affected families.

As part of an ongoing international collaboration, genomic studies were undertaken in a number of OCA families from several communities in Pakistan to define the specific molecular causes of disease. In parallel with this, an extensive literature review of previous reported genetic causes of OCA in individuals from Pakistan was undertaken. This knowledge advances the understanding of the relative contribution of pathogenic OCA variants in Pakistan.

3.3.2 Materials and methods

This study entails the genetic and clinical investigations in 36 Pakistani families, from different ethnic groups and provinces in Pakistan (Table 3.5). Medical histories were taken from all families, and a diagnosis of OCA was established in all affected individuals. Following results from genetic analyses and the identification of pathogenic variants in OCA-associated disease genes, affected members in all families were revisited, and a detailed medical history was ascertained. Facial photographs and videos were taken with consent to document phenotypic features and confirm disease status. Ophthalmic examinations were completed using the best locally available resources with support from visiting ophthalmologists from local collaborating hospitals. This included visual acuity testing using Snellen charts and LVRC (low vision

resource centre) numbers – LVRC distance LogMAR Visual Acuity Chart, colour vision testing using Ishihara charts and funduscopic examination by direct ophthalmoscopy where possible in the affected individuals examined.

Blood samples were taken with informed consent for DNA extraction (see section 2.3.2). As the *TYR* gene is a small gene comprising only five coding exons, all families were first investigated by targeted dideoxy sequencing for variants in one affected individual in each family, using primers designed as described in section 2.3.2 (Appendix Table D2). This defined the likely disease-causing variant(s) in 21 families (families 5 - 25).

In all remaining families (families 26 - 40), next generation sequencing was performed on a single affected individual in each family using the Illumina TruSight™ One clinical exome sequencing panel as described in section 2.3.5. TruSight™ One clinical exome sequencing was also subsequently performed in a single affected individual (IV:7) in family 23, and in a molecularly undiagnosed affected individual (VI:5) in family 33. Bioinformatic analysis of exome data was performed as per section 2.3.5. Primer design (Appendix Table D2), PCR and dideoxy sequencing was performed as described in section 2.3.3. to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals. SNP genotyping was performed in two affected individuals in family 40 (IV:2 and IV:3) as described in section 2.3.4.

A literature review was performed as described in section 2.4 to retrieve all variants and loci associated with OCA in Pakistan. Findings are summarised in Appendix C.

3.3.3 Results: clinical and genetic findings

Study subjects from 36 families with OCA were enrolled from the Punjab, KPK and Balochistan provinces in Pakistan. The apparent mode of inheritance in all families was consistent with an autosomal recessive disorder. All affected individuals displayed the cardinal clinical features of OCA with white-to-golden

blonde hair, pale-to-reddish white skin, decreased visual acuity, nystagmus and photophobia. Clinical examination in all affected individuals examined demonstrated the classical ophthalmic features of albinism, namely: nystagmus, foveal hypoplasia and a hypopigmented albinotic fundus.

Targeted dideoxy sequencing of the *TYR* gene identified a number of novel and previously reported variants in families 5 - 25 (findings summarised in Table 3.5). All identified variants appeared to segregate appropriately for autosomal recessive disease in all families investigated (Figure 3.4), apart from family 23, where a well reported *TYR* NM_000372.4:c.1037-7T>A splice variant (153, 184-190) was initially identified through targeted *TYR* gene sequencing in individual IV:5 in homozygous form. This variant segregated appropriately in affected and unaffected individuals in family 23 apart from affected individual IV:7, in whom the *TYR* c.1037-7T>A variant was only detected in heterozygous form. TruSight™ One clinical exome sequencing was subsequently performed in this individual, which led to the identification of a further heterozygous *TYR* variant, c.1255G>A; p.(Gly419Arg) that has been reported in numerous Pakistani families residing in several regions of Pakistan including Punjab, KPK, Sindh, and Azad Jammu and Kashmir (131, 186, 188, 191-194) (Table 3.5). This variant was subsequently confirmed by dideoxy sequencing in individual IV:7, but was not present in affected individuals IV:3 and IV:5.

TruSight™ One clinical exome sequencing identified a number of novel and known variants in *OCA2* in families 26 - 39, with findings summarised in Table 3.5. All identified variants segregated appropriately for an autosomal recessive disease in all families investigated (Figure 3.4), apart from family 33, where the previously reported *OCA2* NM_000275.3:c.1045-15T>G splice variant (186, 193) was identified in homozygous form in affected individuals IV:1, IV:5, V:3, V:8, VI:2 and VI:3, but only in heterozygous form in affected individual VI:5. TruSight™ One clinical exome sequencing was subsequently performed in this individual, but this did not identify any further candidate disease variants in *OCA2* or other OCA associated disease genes.

The novel *TYR* c.132T>A; p.(Ser44Arg) and c.1002delA; p.(Ala335Leufs*20) variants were present in the gnomAD population database in heterozygous form in 5 individuals from the Latino/admixed American population (gnomAD v2.1.1; MAF 0.00001989) and 1 individual from the South Asian population (gnomAD v3.1.1; MAF of 0.000006571), with no homozygous individuals identified. The novel *OCA2* c.1762C>T; p.(Arg588Trp) variant was present in homozygous form in 5 South Asian individuals in gnomAD (v2.1.1), with a MAF of 0.0009973 (MAF of 0.005131 in the South Asian population). The remaining novel variants identified in this study, *TYR* c.240G>C; p.(Trp80Cys), *TYR* c.667C>T; p.(Gln223*) and *OCA2* c.2458T>C; p.(Ser820Pro) were all absent in gnomAD (Table 3.6). *In silico* analysis of the novel variants identified was undertaken using the variant prediction tools SIFT, PolyPhen-2, PROVEAN and MutationTaster2. These were largely in agreement and predicted a deleterious effect for all novel variants assessed, apart from *OCA2* c.1762C>T; p.(Arg588Trp), where pathogenicity predictions were conflicting (Table 3.6). Localisation of the novel *TYR* and *OCA2* missense variants within the respective polypeptide is shown in Figure 3.3 alongside conservation analysis of the affected amino acid residue across a range of vertebrate species.

In family 40, TruSight™ One clinical exome sequencing in a single affected individual initially failed to identify any candidate disease variants likely responsible for the OCA phenotype. Whole genome SNP mapping was subsequently performed in both affected individuals, and identified a 7.42 Mb region of homozygosity in chromosome 15q13.1 (GRCh38 chr15:g.22180552; rs4462663 to chr15:g.29601735; rs1459361) common to both affected family members and encompassing the *OCA2* gene (Figure 3.5B). Array analysis identified an intragenic homozygous deletion of ~63.4 kb within the *OCA2* gene in both affected individuals, resulting in a deletion of exon 19, and predicted to result in a frameshift (Figure 3.5B). Exome data, analysed visually using IGV, indicated no reads over exon 19, with a good depth of sequencing reads covering the adjacent exons 18 and 20, again supporting a homozygous intragenic deletion (Figure 3.5C).

A literature and genomic database review of known causes of OCA in families from Pakistan was undertaken. Together with the current study, this identified reports of 838 individuals from 197 Pakistani families with 93 candidate mapped loci or causative genetic variants in 13 genes or loci (Appendix C), with variable levels of confidence of causality.

Table 3.5 *TYR* and *OCA2* variants segregating with albinism identified in this study

Gene	Nucleotide variant	Protein variant	References	gnomAD v2.1.1 MAF all/SAS (hom count)	ClinVar (Accession)	Family	Province (region)	Caste
TYR NM_000372.5	c.132T>A	p.(Ser44Arg)	Novel	0.00001989/ 0 (0)	Not present	5	Balochistan	Khatak
						6	KPK (Peshawar)	Pashton
	c.240G>C	p.(Trp80Cys)	Novel	Not present	Not present	7	KPK	Pathan
						8	KPK	Pathan
	c.272G>A	p.(Cys91Tyr)	Chong *et al* (195)	0.000003982/ 0.00003266 (0)	Pathogenic (VCV000039977)	9	Punjab (RYK)	Somro
	c.308G>A	p.(Cys103Tyr)	Lasseaux *et al* (141)	Not present	Not present	10	Punjab (Sahiwal)	Khokar
	c.346C>T	p.(Arg116*)	Oetting[a] *et al* (196); Wei *et al* (162); Gul[a] *et al* (192); Thomas *et al* (144); Lin *et al* (197); Zhong *et al* (189)	0.00002829/ 0.00003266 (0)	Pathogenic (VCV000099565)	11	Punjab (Lahore)	Mughal
	c.649C>T	p.(Arg217Trp)	Tripathi *et al* (198); Shahzad *et al* (186)	0.0001914/ 0.0003928 (0)	Pathogenic/ Likely pathogenic/ VUS (VCV000003795)	12	Punjab	Rajput
	c.667C>T	p.(Gln223*)	Novel	Not present	Not present	13	KPK (Bunar)	Pashton
	c.715C>T	p.(Arg239Trp)	Nakamura *et al* (199); Renugadevi *et al* (200); Zhong *et al* (189)	0.00002790/ 0.00006551 (0)	Not present	14	Punjab (RYK)	Turk
	c.832C>T	p.(Arg278*)	Tripathi *et al* (194); Takizawa *et al* (201); Wei[a] *et al* (162); Wei[b] *et al* (171); Wang *et al* (202); Shahzad *et al* (186); Lionel *et al* (203); Rego *et al* (204); Lin *et al* (197); Zhong *et al* (189); Bibi *et al* (205)	0.0001699/ 0.001274 (0)	Pathogenic (VCV000099583)	15	Punjab	Punjabi
						16	Punjab (Lahore)	Gujjar
						17	KPK (Peshawar)	Pashton

	c.1255G>A	p.(Gly419Arg)	King *et al* (206); Chaki *et al* (207); Gul[a] *et al* (192); Shahzad *et al* (186); Gul[b] *et al* (188)	0.00006032/ 0.0003921 (0)	Pathogenic (VCV000003792)	18	Punjab (Chiniot)	Chudhar Jutt
						19	Punjab (Gujranwala)	Arain
						20	Punjab	Virk
						21	Punjab (Lahore)	Malik Awan
						22	NA	NA
	c.1037-7T>A	Affects splicing	Spritz[a] *et al* (184); Ribero *et al* (185); Shahzad *et al* (186); Marti *et al* (153); Ceyhan-Birsoy *et al* (187); Gul[b] *et al* (188); Zhong *et al* (189); Hou *et al* (190)	0.0008614/ 0.00006551 (1)	Pathogenic/ Likely pathogenic (VCV000099527)	23	Punjab (Sargodha)	Cheema
	c.1255G>A	p.(Gly419Arg)	See above entry for families 18-22	See above entry for families 18-22	See above entry for families 18-22			
	c.832C>T	p.(Arg278*)	See above entry for families 15-17	See above entry for families 15-17	See above entry for families 15-17	24	Punjab (Lahore)	Mayo
	c.1002delA	p.(Ala335Leu fs*20)	Novel	Not present	Not present			
	c.832C>T	p.(Arg278*)	See above entry for families 15-17	See above entry for families 15-17	See above entry for families 15-17	25	KPK (Peshawar)	Pashton
	c.1255G>A	p.(Gly419Arg)	See above entry for families 18-22	See above entry for families 18-22	See above entry for families 18-22			
OCA2 NM_000275.3	c.1456G>T	p.(Asp486Tyr)	Jaworek *et al* (193); Shahzad *et al* (186)	0.00002386/ 0.0001960 (0)	VUS (VCV000617806)	26	Punjab (Qasur)	Dogar
						27	Punjab (Borewala)	Dogar
						28	NA	NA
	c.2207C>T	p.(Ser736Leu)	Spritz[b] *et al* (208); Marti *et al* (153);Yang *et al* (209)	0.00002387/ 0 (0)	Likely pathogenic/ VUS (VCV000195557)	29	NA	NA

c.2360C>T	p.(Ala787Val)	Oetting[b] *et al* (210); Zhang *et al* (211)	0.00002475/ 0 (0)	Pathogenic/ likely pathogenic (VCV000617810)	30	KPK (Bajaur)	NA
c.2458T>C	p.(Ser820Pro)	Novel	Not present	Not present	31	Punjab	Niaz
c.1045-15T>G	Affects splicing	Jaworek *et al* (193); Shahzad *et al* (186)	0.00002394/ 0.0001960 (0)	Likely pathogenic (VCV000617802)	32	Punjab	Niaz
					33	KPK	Afridi
					34	KPK (Peshawar)	Pashton
					35	KPK (Peshawar)	Pashton
					36	KPK (Peshawar)	Pashton
c.408_409delAA	p.(Arg137Ilefs* 83)	Novel	Not present	Not present	37	KPK (Peshawar)	Pashton
c.1327G>A	p.(Val443Ile)	Lee *et al* (212); Norman *et al* (213); Marti *et al* (153); Rego *et al* (204); Campbell *et al* (214); Hou *et al* (190)	0.003056/ 0.0001307 (4)	Pathogenic/ VUS (VCV000000955)	38	Punjab	Saraiki
c.1762C>T	p.(Arg588Trp)	Novel	0.0009973/ 0.005131 (5)	VUS (VCV000885235)			
c.2020C>G	p.(Leu674Val)	Mondal *et al* (215); Norman *et al* (213)	0.0003223/ 0.002580 (1)	Pathogenic/ VUS (VCV000194918)	39	KPK	Yousafzai
c.408_409delAA	p.(Arg137Ilefs* 83)	See above entry for family 37	See above entry for family 37	See above entry for family 37			
Exon 19 deletion	Frameshift	-	-	-	40	NA	NA

Abbreviations: gnomAD, genome aggregation database; hom, homozygous; KPK, Khyber Pakhtunkhwa; MAF, minor allele frequency; RYK, Rahim Yar Khan; VUS, variant of uncertain significance. Novel variants are highlighted in red.

Table 3.6 Novel *TYR* and *OCA2* variants identified in this study

Gene	Variant	gnomAD v2.1.1			*In silico*			
		MAF (all)	MAF (SAS)	Hom count	SIFT	PolyPhen-2	Provean	MutationTaster2
TYR NM_000372.5	c.132T>A; p.(Ser44Arg)	0.00001989	0	0	Deleterious	Possibly damaging	Deleterious	Disease causing
	c.240G>C; p.(Trp80Cys)	Not present	Not present	Not present	Deleterious	Probably damaging	Deleterious	Disease causing
	c.667C>T; p.(Gln223*)	Not present	Not present	Not present	-	-	-	Disease causing
	c.1002delA; p.(Ala335Leufs*20)	Not present	Not present	Not present	-	-	-	Disease causing
OCA2 NM_000275.3	c.2458T>C; p.(Ser820Pro)	Not present	Not present	Not present	Deleterious	Possibly damaging	Deleterious	Disease causing
	c.1762C>T; p.(Arg588Trp)	0.0009973	0.005131	5	Deleterious	Benign	Deleterious	Polymorphism
	c.408_409delAA; p.(Arg137Ilefs*83)	Not present	Not present	Not present	-	-	-	Disease causing

Abbreviations: gnomAD, genome aggregation database; Hom, homozygous; MAF, minor allele frequency, SAS, South Asian. SIFT, Polyphen-2 and Provean in silico predictions unavailable for nonsense and frameshift variants.

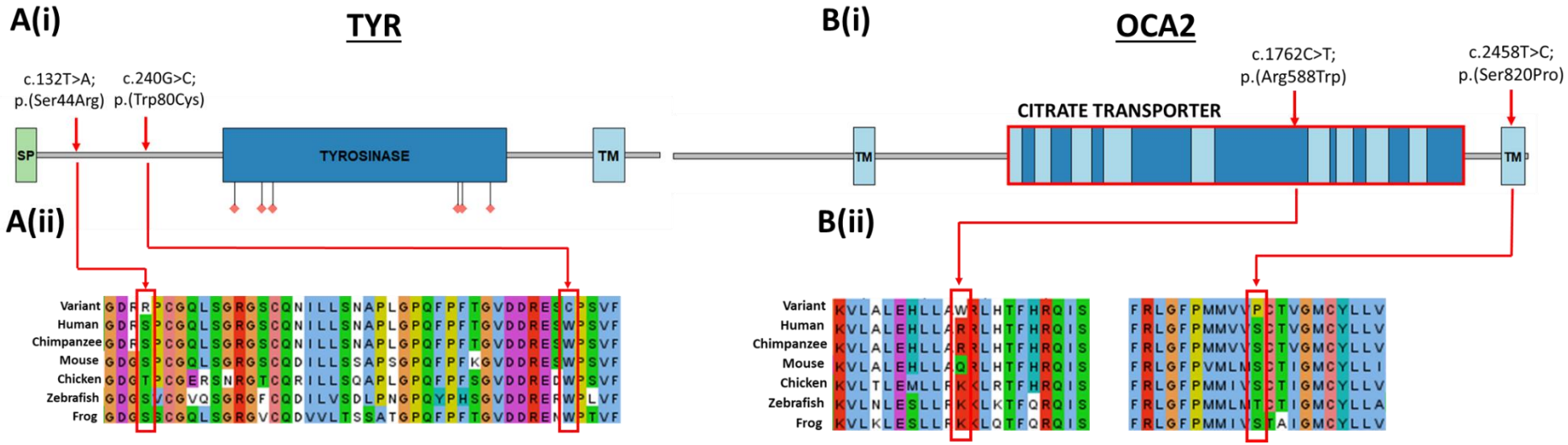
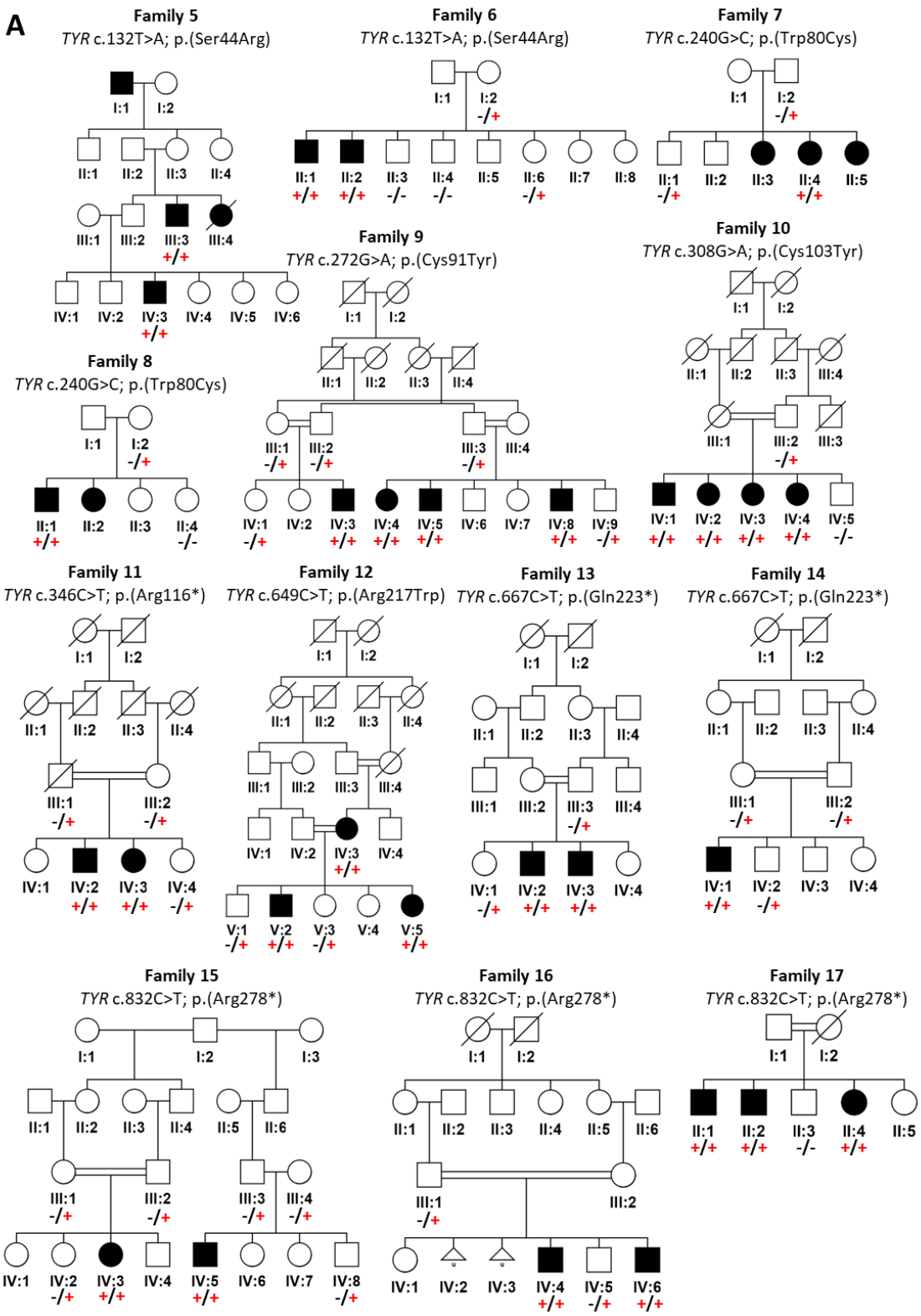


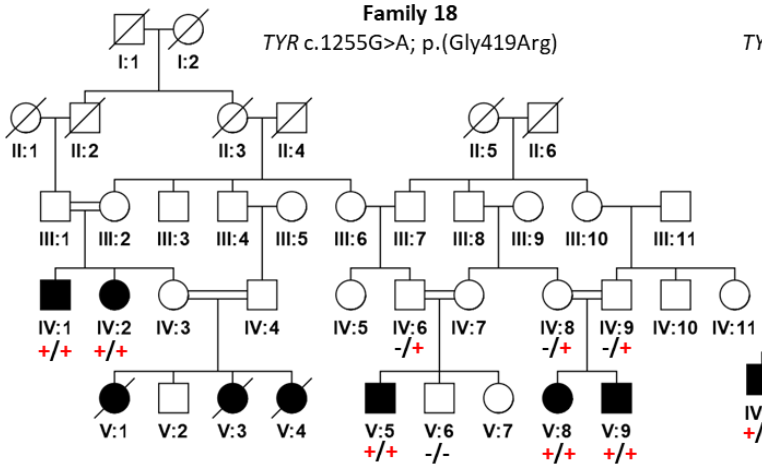
Figure 3.3 Novel missense *TYR* and *OCA2* variants identified in this study

Schematic showing domain architecture of A(i) TYR and B(i) OCA2 [adapted from UniProt (216)] and the location of novel missense variants identified. Abbreviations; SP, signal peptide; TM (in light blue), transmembrane domain. For the TYR polypeptide, the red diamonds denote the histidine residues that bind to copper atoms and hence structurally coordinate the positions of the copper-containing catalytic binding sites. For the OCA2 polypeptide, the citrate transporter domain is outlined in red. Conservation analysis: multiple species alignments of amino acid sequences of A(ii) TYR *and B(ii)* OCA2 *at variant locations*


A
Family 5
TYR c.132T>A; p.(Ser44Arg)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 III:2 III:3 +/+ III:4
IV:1 IV:2 IV:3 +/+ IV:4 IV:5 IV:6
Family 6
TYR c.132T>A; p.(Ser44Arg)
I:1 I:2 -/+
II:1 +/+ II:2 +/+ II:3 -/- II:4 -/- II:5 II:6 -/+ II:7 II:8
Family 7
TYR c.240G>C; p.(Trp80Cys)
I:1 I:2 -/+
II:1 -/+ II:2 II:3 II:4 +/+ II:5
Family 8
TYR c.240G>C; p.(Trp80Cys)
I:1 I:2 -/+
II:1 +/+ II:2 II:3 II:4 -/-
Family 9
TYR c.272G>A; p.(Cys91Tyr)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 -/+ III:2 -/+ III:3 -/+ III:4
IV:1 -/+ IV:2 IV:3 +/+ IV:4 +/+ IV:5 +/+ IV:6 IV:7 IV:8 +/+ IV:9 -/+
Family 10
TYR c.308G>A; p.(Cys103Tyr)
I:1 I:2
II:1 II:2 II:3 III:4
III:1 III:2 -/+ III:3
IV:1 +/+ IV:2 +/+ IV:3 +/+ IV:4 +/+ IV:5 -/-
Family 11
TYR c.346C>T; p.(Arg116*)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 -/+ III:2 -/+
IV:1 IV:2 +/+ IV:3 +/+ IV:4 -/+
Family 12
TYR c.649C>T; p.(Arg217Trp)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 III:2 III:3 III:4
IV:1 IV:2 IV:3 +/+ IV:4
V:1 -/+ V:2 +/+ V:3 -/+ V:4 V:5 +/+
Family 13
TYR c.667C>T; p.(Gln223*)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 III:2 III:3 -/+ III:4
IV:1 -/+ IV:2 +/+ IV:3 +/+ IV:4
Family 14
TYR c.667C>T; p.(Gln223*)
I:1 I:2
II:1 II:2 II:3 II:4
III:1 -/+ III:2 -/+
IV:1 +/+ IV:2 -/+ IV:3 IV:4
Family 15
TYR c.832C>T; p.(Arg278*)
I:1 I:2 I:3
II:1 II:2 II:3 II:4 II:5 II:6
III:1 -/+ III:2 -/+ III:3 -/+ III:4 -/+
IV:1 IV:2 -/+ IV:3 +/+ IV:4 IV:5 +/+ IV:6 IV:7 IV:8 -/+
Family 16
TYR c.832C>T; p.(Arg278*)
I:1 I:2
II:1 II:2 II:3 II:4 II:5 II:6
III:1 -/+ III:2
IV:1 IV:2 IV:3 IV:4 +/+ IV:5 -/+ IV:6 +/+
Family 17
TYR c.832C>T; p.(Arg278*)
I:1 I:2
II:1 +/+ II:2 +/+ II:3 -/- II:4 +/+ II:5

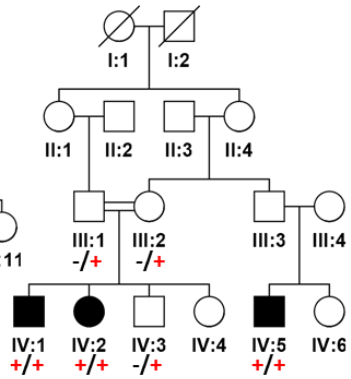
Family 18

TYR c.1255G>A; p.(Gly419Arg)



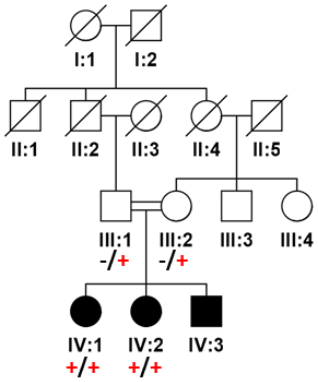
Family 19

TYR c.1255G>A; p.(Gly419Arg)



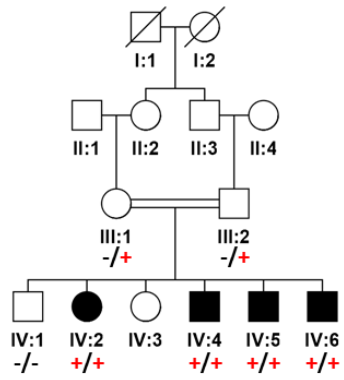
Family 20

TYR c.1255G>A; p.(Gly419Arg)



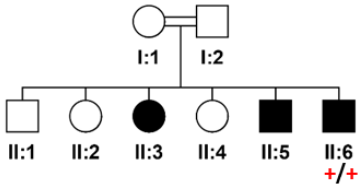
Family 21

TYR c.1255G>A; p.(Gly419Arg)



Family 22

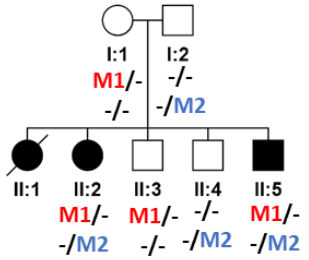
TYR c.1255G>A; p.(Gly419Arg)



Family 24

M1: *TYR* c.832C>T; p.(Arg278*)

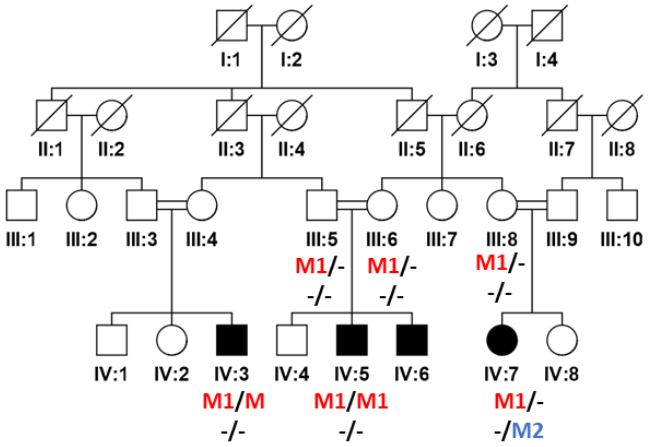
M2: *TYR* c.1002delA; p.(Ala335Leufs*20)



Family 23

M1: *TYR* c.1037-7T>A

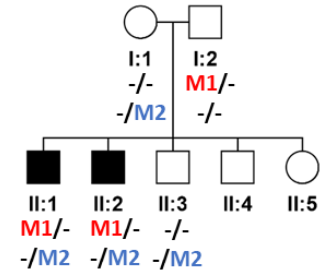
M2: *TYR* c.1255G>A; p.(Gly419Arg)



Family 25

M1: *TYR* c.832C>T; p.(Arg278*)

M2: *TYR* c.1255G>A; p.(Gly419Arg)



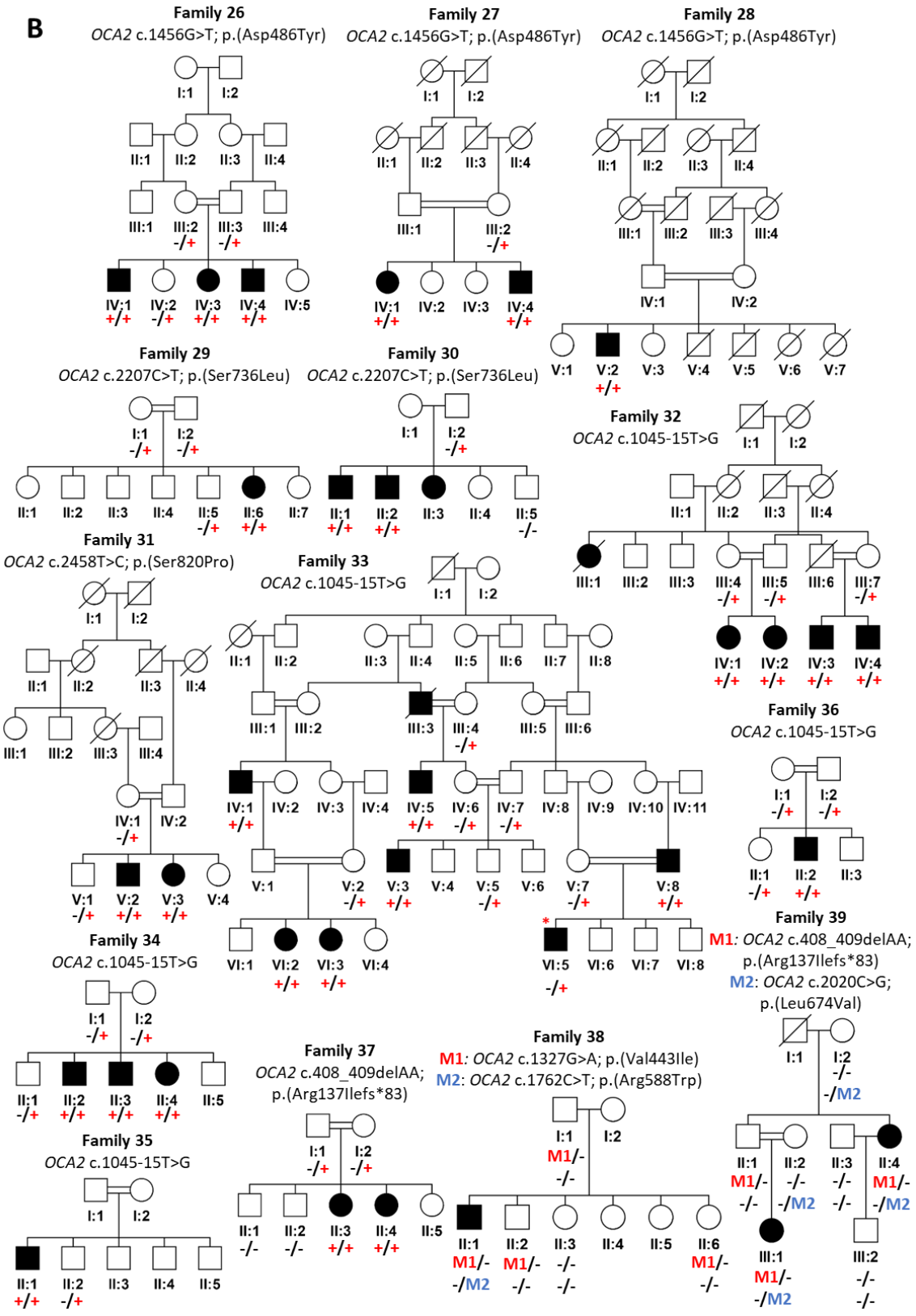


Figure 3.4 Pedigrees and genotype data for families 5 - 39

Pedigree diagrams of families 5 - 39 investigated with OCA, showing segregation of (A) TYR *and (B)* OCA2 *variants identified. Genotypes are shown beneath each family member investigated (+, variant; -, wild type).*

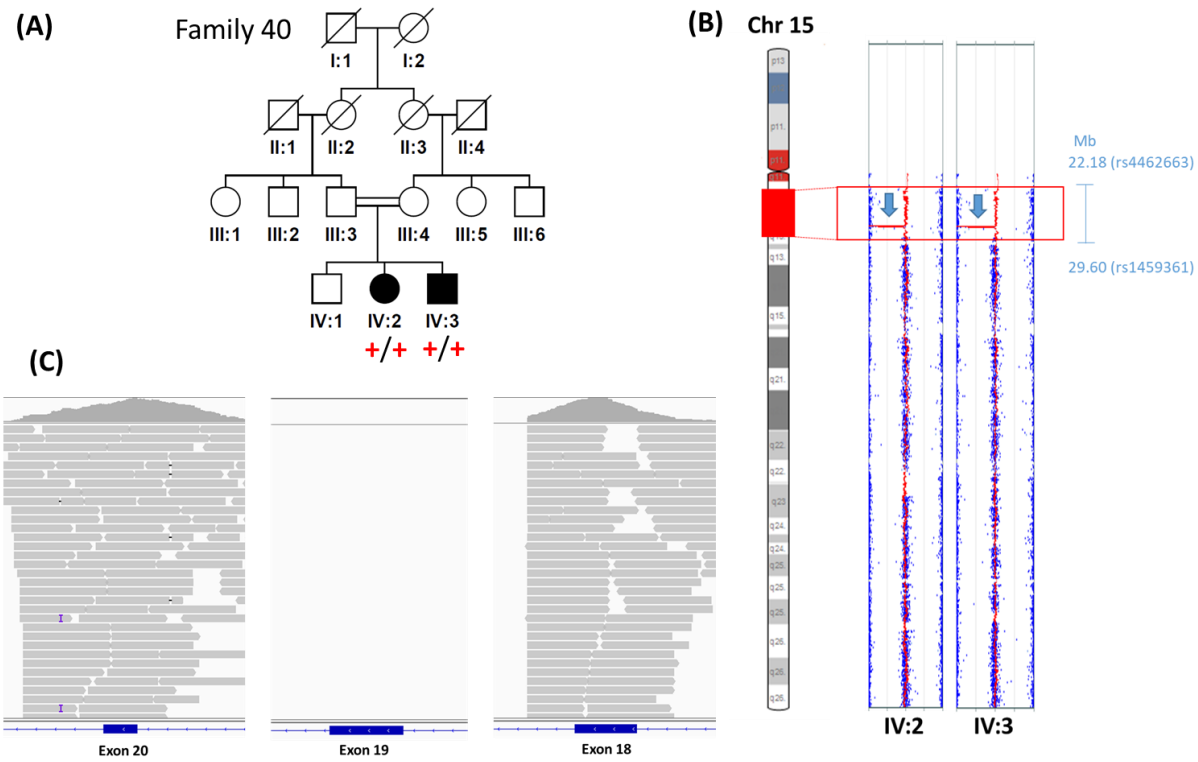


Figure 3.5 Family 40 pedigree and *OCA2* genotype data

(A) Pedigree of family 40 showing segregation of OCA2 *structural variant identified. Genotypes are shown beneath each family member investigated (+, variant; -, wild type). (B) Screenshot from KaryoStudio software showing ideogram of chromosome 15 and the loss-of-heterozygosity region encompassing* OCA2*, shared between both affected individuals IV:2 and IV:3, as well as a homozygous deletion within this region in the* OCA2 *gene (C) Integrative Genome Viewer (IGV) screenshot showing absence of sequence reads over exon 19 of* OCA2*, with good read depth and coverage over the adjacent exons 18 and 20*

3.3.4 Discussion

This study describes the findings from genetic analyses of 36 families of Pakistani descent, identifying a total of 22 likely disease-associated variants in *TYR* and *OCA2* responsible for OCA in Pakistani communities. These studies have identified four novel *TYR* variants and three novel *OCA2* variants (Table 3.6), expanding the molecular spectrum of disease-causing OCA variants globally.

Four of the OCA-associated variants identified in this study, *TYR* c.1037-7T>A (family 23), and *OCA2* p.(Leu674Val) (family 39), p.(Val443Ile) and p.(Arg588Trp) (both family 38), are present in gnomAD in homozygous form in one, one, four and five individuals respectively. The individual homozygous for *TYR* c.1037-7T>A is of Ashkenazi Jewish origin, whilst the individuals in gnomAD homozygous for the above

OCA2 variants are all of South Asian ancestry. These variants all segregated with disease in compound heterozygous form in their respective families. Hypomorphic *TYR* variants are well defined in the aetiology of the OCA1B subtype, where reduced tyrosinase activity results in a milder phenotype with reduced levels of pigmentation in affected individuals. These variants may therefore represent hypomorphic variants that cause only a partial loss of *TYR* or *OCA2* gene function, with homozygotes exhibiting a mild phenotype, accounting for the homozygous individuals present in the gnomAD database.

The *OCA2* p.(Leu674Val) variant has previously been described in association with OCA in two Indian individuals in both homozygous and compound heterozygous form (215). Both individuals exhibited an incomplete albinism phenotype, with the individual heterozygous for the p.(Leu674Val) as well as a c.775dupG variant showing clinical features of nystagmus, hazel irides, light golden-brown hair and pinkish skin. The individual homozygous for the p.(Leu674Val) variant showed clinical features of brown irides with iris transillumination, silky-brown hair and very fair pinkish skin with no apparent nystagmus (215). This study detected the same p.(Leu674Val) variant in compound heterozygous form in family 39, together with a novel frameshift *OCA2* variant, with affected individuals again displaying an incomplete OCA phenotype with nystagmus, blue irides, light brown hair and pink skin. The p.(Leu674Val) variant may therefore represent a milder *OCA2* mutation contributing to an incomplete OCA phenotype when occurring in conjunction in compound heterozygous form with a more deleterious *OCA2* variant.

Although individuals homozygous for the *TYR* c.1037-7T>A and *OCA2* p.(Val443Ile) variants have been identified in gnomAD, these variants have also previously been described in multiple OCA individuals in European and Chinese populations (153, 184, 189, 211, 212, 217), supporting pathogenicity. Some affected individuals compound heterozygous for the *OCA2* p.(Val443Ile) variant have been described as having a partial albinism phenotype (213, 214), suggesting that these may be pathogenic, albeit hypomorphic, variants. The *OCA2* p.(Arg588Trp) variant affects a moderately conserved amino acid residue, although multiple *in silico* prediction tools are conflicting in their predictions of pathogenicity (Table 3.6). For affected individual II:1 in family 38 who was compound heterozygous for *OCA2* p.(Arg588Trp) and

p.(Val443Ile) variants, re-analysis of the exome data could not identify any other candidate variants in any of the known OCA associated genes. Ultimately, functional characterisation of the variant would be helpful to determine the nature of the mutation and the extent of its biological impairment (218).

In family 23, two different *TYR* variants [c.1037-7T>A and p.(Gly419Arg)] were identified segregating with OCA within the same extended consanguineous family. The presence of >1 disease-associated variant in a family may give rise to atypical inheritance patterns or phenotypical outcomes, complicating interpretation of linkage analysis and co-segregation results, clinical interpretation, patient counselling and management. This intra-familial locus heterogeneity is particularly relevant within community isolates including those within Pakistan, and has been demonstrated to occur in up to 15% of a cohort of Pakistani families with presumed autosomal recessive hearing loss (219). In consanguineous families with likely monogenic diseases, where pedigree analysis suggests an autosomal recessive inheritance, but where there is failure to identify a single rare segregating candidate disease variant in all affected individuals, the possibility of a second disease variant or even a second disease gene should be considered.

In family 33, the previously described disease-associated *OCA2* c.1045-15T>G splice variant was detected in six out of seven affected individuals, but only in heterozygous form in the seventh affected individual (individual VI:5, highlighted by an (*) in Figure 3.4), with no other candidate variants in any of the known OCA associated genes detected in this individual. Interestingly, this heterozygous *OCA2* c.1045-15T>G individual displayed a somewhat different phenotype compared to affected individuals who were homozygous for the variant, with a darker hair colour (golden brown instead of blonde). Unlike in family 23 described above, further genetic studies failed to identify a second variant in the same OCA disease gene responsible for disease in this individual. There is a high level of missing heritability in OCA, with 25% of patients investigated only having detectable mutations in a single OCA allele (129). This may be due to variants in the gene promoter or other regulatory regions that may not be detected or recognised by current sequencing technologies. It is known that the phenotype of OCA2 can be modified by *MC1R* or *TYRP1* mutations, demonstrating a synergistic interaction between genes throughout this pigment pathway (220, 221).

Possible digenic inheritance involving genetic interactions between heterozygous *TYR* and *OCA2* variants has also recently been described in three Pakistani OCA families (191). Consequently, the missing heritability in this individual may reflect an undetected mutation in an OCA disease gene or other pigment pathway gene that interacts with *OCA2*. Interestingly, our research group has recently provided strong supporting evidence for pathogenicity of a *TYR* haplotype consisting of a combination of two common hypomorphic *TYR* variants in *cis*, which when inherited in *trans* to a known *TYR* deleterious mutation, is likely to account for a proportion of OCA1B cases with apparent missing heritability (see chapter 3.2). A similar situation may explain the OCA2 missing heritability in family 33.

Gross deletions in the *OCA2* gene are increasingly being recognised as a molecular mechanism of disease in OCA2, with 42 such mutations reported to date (HGMD; accessed 15.04.2021), including the 2.7 kb intragenic deletion spanning exon 7 that is the most common cause of OCA2 in individuals of sub-Saharan African heritage (222). Intragenic *OCA2* deletions are also increasingly recognised in Pakistani communities (140, 186, 188). Shahzad *et al* reported five gross deletions in *OCA2*, including one family of Malick ethnicity with an exon 19 deletion (186). More recently, a further *OCA2* exon 19 deletion was reported in two further Pakistani families of Pashtun origin recruited from Tank city, KPK, with different deletion coordinates to the family reported by Shahzad *et al* (188). One limitation of this study is the inability to accurately map breakpoints with the genomic strategies employed, and therefore it is not possible to determine if the *OCA2* exon 19 deletion in family 40 is the same as either of those previously reported in Pakistani families. Prior studies do however support *OCA2* exon 19 deletion as a mechanism of disease, with the exon skipping predicted to result in a frameshift and loss of function via nonsense-mediated decay. The significant contribution of *OCA2* gross deletions to OCA in Pakistani families also supports the need for genomic or bioinformatic strategies for their detection within this community.

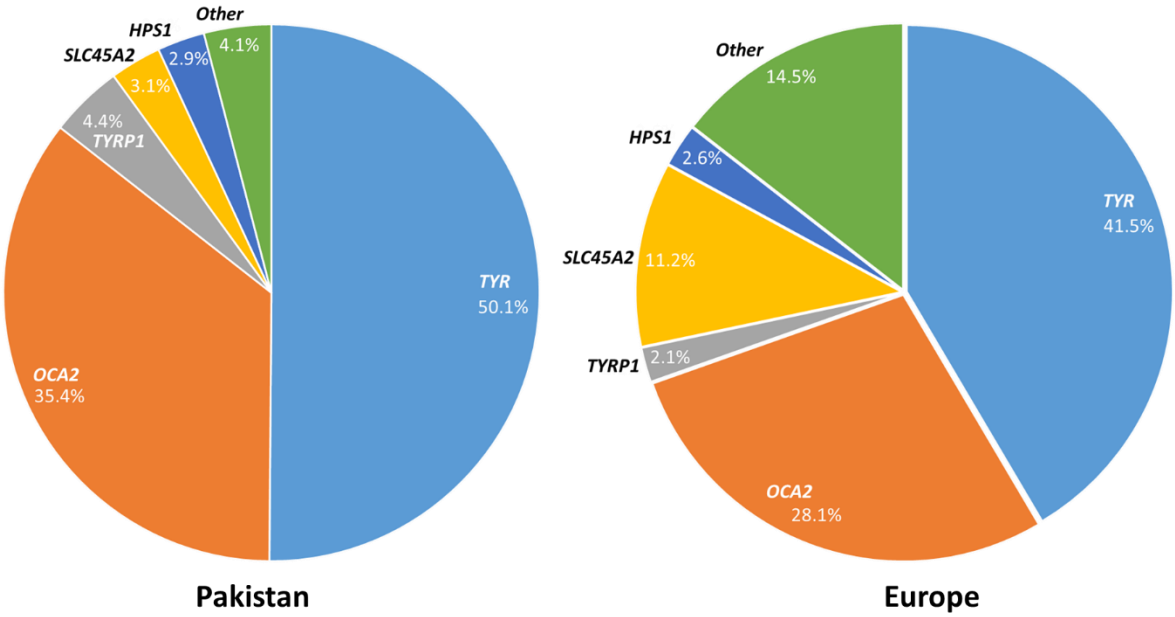


Figure 3.6 Contribution of OCA genes to OCA within Pakistan and Europe

European figures are derived from (141, 153)

The genetic studies described here, together with a comprehensive literature review encompassing a combined cohort of 838 individuals from 197 Pakistani families (Appendix C), highlight *TYR* and *OCA2* as the major genes contributing to OCA in Pakistan comparable to the prevalence in European populations (141, 153), accounting for 50.1% and 35.4% of molecularly diagnosed cases respectively (Figure 3.6).

While most of the *TYR* and *OCA2* variants described in this study are present in only one or a few families, some appear to be commonly associated with OCA in Pakistan. These include *TYR* variants p.(Arg278*) (identified in families 15 - 17, 24, and 25) and p.(Gly419Arg) (identified in families 18-22, 23, 25), as well as *OCA2* variants c.1045-15T>G (identified in families 32 - 36) and p.(Asp486Tyr) (identified in families 26 - 28), which each account for ~20% of all *TYR* or *OCA2* alleles in Pakistani OCA families (Appendix C). While it is not possible to conclusively determine whether common variants represent recurrent ('hot spot') gene variants or regional founder variants without more detailed genetic analyses, there is evidence in the literature to support both mechanisms. For example, the *OCA2* c.1045-15T>G and p.(Asp486Tyr) variants have not been reported outside Pakistan, and most likely stem from a single gene mutation event that occurred in a founder ancestor individual and which was

transmitted to subsequent generations, accumulating at an increased frequency in the community. Conversely, whilst the *TYR* p.(Arg278*) variant appears to occur at an increased frequency in the Pakistani population, this variant has also been identified in numerous OCA families from various ethnicities and geographical locations, including India (131, 215, 223-225), Guyana (194), Japan (201, 226, 227), China (162, 171, 189, 197, 202, 228-230), Korea (231), Europe (131), Mexico (131) and Israel (232), and likely represents a recurrent gene mutation.

Founder mutations often represent important causes of disease in a particular region, and knowledge of their presence, frequency and clinical outcomes is of enormous value for local healthcare resource planning and for designing community-specific diagnostic testing and counselling protocols. As with ancestral founder mutations, knowledge of recurrent gene mutations is of great utility in the development of cost-effective disease-specific genetic testing strategies.

3.4 Conclusions and future work

This chapter details the significant scientific and clinical insights gained from studies of OCA in families from Amish and Pakistani communities. The unique genomic architecture of the Amish community, together with the relatively high frequency of OCA type 1B in the population, enabled empowered genetic studies able to determine the haplotype, phasing and inheritance of the hypomorphic p.(Ser192Tyr) and p.(Arg402Gln) *TYR* variants in a large number of related individuals for the first time. The pathogenicity of these two common *TYR* variants has been heavily debated for some time, and this study now provides irrefutably strong evidence that the *TYR* p.(Ser192Tyr) and p.(Arg402Gln) variants are pathogenic when inherited in *cis*. This finding has important diagnostic implications, as considering and reporting this in-*cis* haplotype as a pathogenic allele could increase the molecular diagnoses in the diagnostically challenging albinism patient cohorts with missing heritability and/or partial or mild albinism phenotypes. It will be crucially important to accurately determine the phase of these common variants, and due to the high frequencies of these variants alone in the population which can limit informative phase studies in relatives, consideration should perhaps be given to the use of amplicon-based long-read sequencing technologies that allow haplotype phasing in the genomic workup of such patients (233).

Studies of families with OCA in Pakistan have identified novel *TYR* and *OCA2* disease variants, and expand current knowledge of the molecular spectrum and specific genetic causes of OCA within Pakistani communities, highlighting a number of regional founder variants. The approach adopted in this study, starting with targeted dideoxy sequencing of *TYR*, followed by exome sequencing in combination with whole genome SNP mapping studies in selected individuals, permits a rapid and cost-effective means of achieving a molecular diagnosis in Pakistani OCA families. In combination with existing datasets, these studies enable accurate genetic testing and provide valuable information to aid the diagnosis and counselling of affected individuals and family members throughout Pakistan and the wider South Asian population.

4 STUDIES OF CILIOPATHIES IN COMMUNITIES

4.1 Introduction

Cilia are evolutionary conserved microscopic, hair-like structures or organelles that protrude from the cell surface. They have important and diverse biological roles in cellular motility and the cell cycle, as well as performing extracellular sensory functions (234). Cilia are broadly divided into two types; motile cilia, and immotile or primary cilia. A single, non-motile (primary) cilium is present in almost every vertebrate cell, whilst cells with motile cilia are present only in specific organs such as the reproductive system or respiratory tract (235). The beating of motile cilia (or flagella) serves to generate fluid movement across cell surfaces, and is responsible for the movement of mucus up the respiratory tract, propulsion of sperm, and establishment of left-right asymmetry in the embryonic node (234). Primary cilia on the other hand transduce extracellular information to the intracellular environment, and have roles in photoreception in photoreceptor cells, olfaction in olfactory neurons, and mechanosensation of fluid flow in kidney epithelial cells (234). Primary cilia also play an important role in several signal transduction pathways, including the noncanonical Wnt/planar cell polarity pathway, and the Hedgehog signalling pathway (234).

The term “ciliopathy” was first used in 1984 to describe the atypical bronchial cilia noted in a subset of children with recurrent respiratory tract infections (236). Today, ciliopathies generally refer to a group of inherited disorders resulting from pathogenic variants in the highly conserved genes involved in the correct assembly and maintenance of cellular cilia or their anchoring structures, the basal bodies, resulting in defective proteins that compromise ciliary structure or function (237, 238). To date, pathogenic variants in approaching 200 genes have been associated with 35 ciliopathy disorders, typically segregating in an autosomal recessive form (239). Ciliopathies arise due to dysfunction of:

- Proteins that primarily localise to - and function within - the ciliary compartment and/or basal body (239). For example, mutations in genes encoding components of the BBsome, an octomeric protein complex functioning as a

cargo adapter and utilising intraflagellar transport machinery for the trafficking of ciliary membrane proteins, result in BBS (240).

- Proteins that are not localised within the cilia, but are required for ciliary formation or function (239). For example, mutations in genes encoding the cytoplasmic proteins required for pre-assembly of the multi-subunit axonemal dynein arms in the cytoplasm prior to their transport into the cilia and flagella, result in primary ciliary dyskinesia (PCD) (241).

Genetic defects associated with ciliopathies can affect both motile and non-motile primary cilia, either separately or in combination (239). Additionally, some ciliary proteins have recognised extraciliary sites and functions (242). These factors likely contribute to the extensive phenotypic heterogeneity associated with ciliopathies (Figure 4.1).

Ciliopathies associated primarily with impairment of motile cilia are characterised by dysfunction of specific tissues and organs that contain motile cilia machinery. The most common motile ciliopathy is PCD, where severe impairment of mucociliary clearance in the respiratory epithelium leads to chronic airway disease. Additionally, defects in ependymal cilia can result in hydrocephalus, defects in sperm flagella or in the cilia in fallopian tubes can lead to subfertility, and cilia defects in the left-right organiser during early embryonic development may result in laterality defects such as situs inversus and heterotaxy (243, 244).

Defects in the sensory and/or signalling functions of primary cilia are often associated with wide phenotypic variability compared to motile ciliopathies. These sensory ciliopathies range from single organ disorders such as *RP1* and *RP1L1*-associated non-syndromic retinal dystrophy, to complex multisystem disorders such as BBS, Joubert syndrome (JBTS) and MORM syndrome (comprising Mental retardation, truncal Obesity, Retinal dystrophy and Micropenis) (239). This likely reflects the varied and widespread cellular and tissue distribution of primary cilia within the human body (245), as well as their key role in many intracellular signal transduction cascades (243). In fact, many of the developmental anomalies associated with syndromic sensory ciliopathies are thought to result from compromised cellular signalling, such as polydactyly (defective Hedgehog signalling) and retinal degeneration (defective G-

protein coupled receptor signalling) seen in BBS and JBTS (246). Retinal degeneration is frequently observed in diverse ciliopathy syndromes (Figure 4.1), caused by dysfunction of the retinal photoreceptor, a highly specialised neuronal cell type whose outer segment compartment is a ciliary organelle of modified structure and function analogous to the primary sensory cilia in other cell types (247).

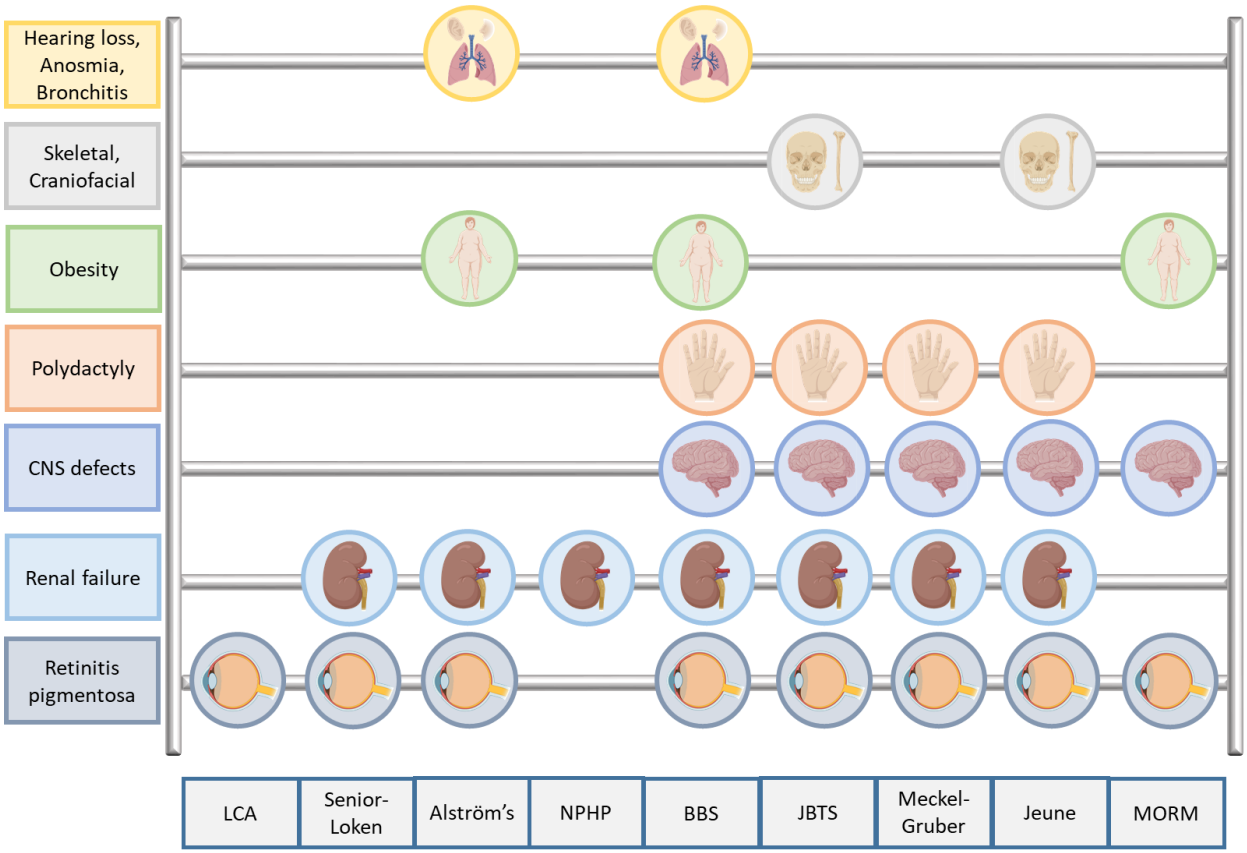


Figure 4.1 Ciliopathy abacus

Schematic representation of common clinical manifestations associated with ciliopathies and their occurrence in each ciliopathy syndrome. The adding of various affected target organs (as an abacus) contributes to the specific ciliopathy syndrome.

Abbreviations: BBS, Bardet-Biedl syndrome; CNS, central nervous system; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; NPHP, nephronophthisis. Image modified from (235, 248) and created with BioRender

This chapter entails a description of clinical and genomic investigations and findings undertaken in Amish and Pakistani families with phenotypic features that are highly suggestive of a ciliopathy disorder. While distinguishing clinical features may be present, these (and other) phenotypically overlapping ciliopathy conditions can be extremely difficult to differentiate clinically. This is particularly the case in developing

nations given the limited resources available for detailed phenotyping, and geographical constraints restricting access to clinical services for affected families. This work was undertaken to broaden the understanding of the molecular basis, nature and spectrum of ciliopathy disorders in communities, and to translate that knowledge into improved regional clinical diagnostic services for these disorders.

Within this chapter, I was responsible for the interpretation and analysis of all collected clinical data for all affected individuals in families 41 – 44. I performed DNA extraction for recruited individuals in family 43 (remaining DNA extraction largely completed by Joe Leslie, University of Exeter). I was also responsible for the analysis of all exome sequencing results, as well as primer design and subsequent cosegregation studies for all variants identified in families within chapters 4.2 and 4.3. I performed the literature review and analysis of all published SCAPER patients described in chapter 4.4, with a particular emphasis on delineating the ocular phenotype.

4.2 Consolidating the phenotypic features of MORM syndrome, and a review of *INPP5E*-related disorders

4.2.1 MORM syndrome

MORM syndrome is an ultra-rare ciliopathy disorder, first described in 2009 in nine affected individuals from a single extended family in Northern Pakistan, caused by homozygosity for a specific nonsense mutation in the *INPP5E* gene (249). In the same year, homozygous missense *INPP5E* variants were identified in seven consanguineous families genetically linked to the first Joubert syndrome locus (JBTS1) on 9q34 (250). JBTS is characterised by a distinctive cerebellar and brainstem malformation (known as the “molar tooth” sign), hypotonia in infancy with later development of ataxia, and developmental delay, and commonly associated with ocular involvement (including progressive retinopathy and/or coloboma) (251-253). The cardinal clinical features of MORM syndrome include intellectual disability (mental retardation), truncal obesity, non-progressive retinal dystrophy and micropenis (251). Both JBTS and MORM syndrome show clinical overlap with ciliopathy syndromes, consolidating *INPP5E* as a ciliopathy disease gene.

Here, clinical and genomic investigations were undertaken in an extended Pakistani family with multiple affected individuals displaying variable phenotypic features suggestive of a ciliopathy disorder. Molecular data and comprehensive clinical assessments, together with a review of all previously published patients, highlight the wide phenotypic spectrum of *INPP5E*-related disorders, and provide an insight into their molecular basis.

4.2.2 Materials and methods

Affected individuals underwent clinical examination at local government hospitals. Blood samples were taken with informed consent for DNA extraction (see section 2.3.2). WES was undertaken using DNA from a single affected individual in family 41 (individual VI:3) at BGI Hong Kong, as described in section 2.3.5. Bioinformatic analysis of exome data was performed as per section 2.3.5, with additional virtual gene panel analysis subsequently undertaken using the “rare multisystem ciliopathy

disorders v1.84" PanelApp Panel (https://panelapp.genomicsengland.co.uk/). Additional filtering was performed to retain heterozygous variants compatible with triallelism (254). Primer design (Appendix Table D2), PCR and dideoxy sequencing was performed as described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals.

A literature review was performed as described in section 2.4 to retrieve all reported *INPP5E* disease-associated variants. Findings are summarised in Table 4.2.

4.2.3 Results: clinical and genetic findings

A large multigenerational extended family (family 41) with nine affected individuals (V:1, VI:2, VI:3, VI:6, VI:7, VI:9 - 12) residing in a remote rural village in Punjab, Pakistan, was investigated. Clinical features in affected individuals included hypotonia, intellectual disability (mild/moderate and non-progressive), obesity, visual impairment, aggressive and hyperactive behaviour, delayed language acquisition, speech impairment and micropenis in males (Table 4.1).

Table 4.1 Summary of clinical features observed in affected individuals in family 41 with MORM syndrome and homozygous for the *INPP5E* p.(Gln627*) variant

Individual	Gender	Age (yrs)	ID	Obesity	Visual Impairment	Hypogonadism	Renal impairment	Post-axial polydactyly
V:1	M	50	✓ (S)	✗	✓	ND	✗	✗
VI:2	M	9	✓ (M)	✗	✓	✓	✗	✗
VI:3	F	7	✓ (M)	✗	✓	✗	✗	✗
VI:6	M	8	✓ (S)	✓	✓	✓	✓	✗
VI:7	M	5	✓ (S)	✓	✓	✓	✓	✗
VI:9	M	10	✓ (S)	✓	✓	✓	✗	✗
VI:10	M	1	✓ (S)	✗	✓	✓	ND	✗
VI:11	M	19	✓ (S)	✗	ND	ND	✗	✗
VI:12	M	15	✓ (S)	✗	ND	✓	✗	✗

Abbreviations: F, female; M, male; ND, no data available; yrs, years; (S), severe/profound; (M), mild/moderate. The (✓) and (✗) symbols indicate the presence of absence of a feature in an affected subject respectively

After filtering WES data in a single affected individual (VI:3), only a single plausible candidate disease variant of potential relevance to the phenotype was identified, entailing a homozygous *INPP5E* nonsense variant (GRCh38) chr9:g.136429731G>A; NM_019892.6:c.1879C>T; p.(Gln627*). This variant was absent in gnomAD (v2.1.1 and v3.1.1) and from a control exome dataset of 100 ethnically matched controls undertaken in the Human Molecular Genetics laboratory in Pakistan. This variant is reported as pathogenic in ClinVar (accession VCV000000396), and has previously been reported in homozygous form in a single family with MORM syndrome from the same province in Pakistan (249, 251). This variant segregated appropriately for an autosomal recessive condition in the family (Figure 4.2).

Additional analysis of exome data undertaken using the “rare multisystem ciliopathy disorders v1.84” PanelApp virtual gene panel did not identify any homozygous or compound heterozygous candidate variants compatible with the phenotype. No heterozygous variants compatible with triallelism were identified.

4.2.4 Discussion

INPP5E, located on chromosome 9q34.3, encodes inositol-polyphosphate 5-phosphatase E, a 72-kDa protein with an N-terminal proline rich domain, a central 5-phosphatase catalytic domain, and a C-terminal CaaX motif. The encoded protein hydrolyses the 5-position phosphate from the inositol ring of the membrane-associated phosphatidylinositols phosphatidylinositol 4,5-bisphosphate [PtdIns(4,5)P_2] and phosphatidylinositol 3,4,5-trisphosphate [PtdIns(3,4,5)P_3] (255). *INPP5E* is widely expressed, with enrichment in the testis and brain (255). In dividing cells, INPP5E localises to the cytosolic face of the Golgi, possibly mediated by its N-terminal proline-rich domain (255). In ciliated cells, the C-terminal CaaX motif is proposed to facilitate localisation of the protein to the primary cilium, where it is concentrated in the axoneme (251).

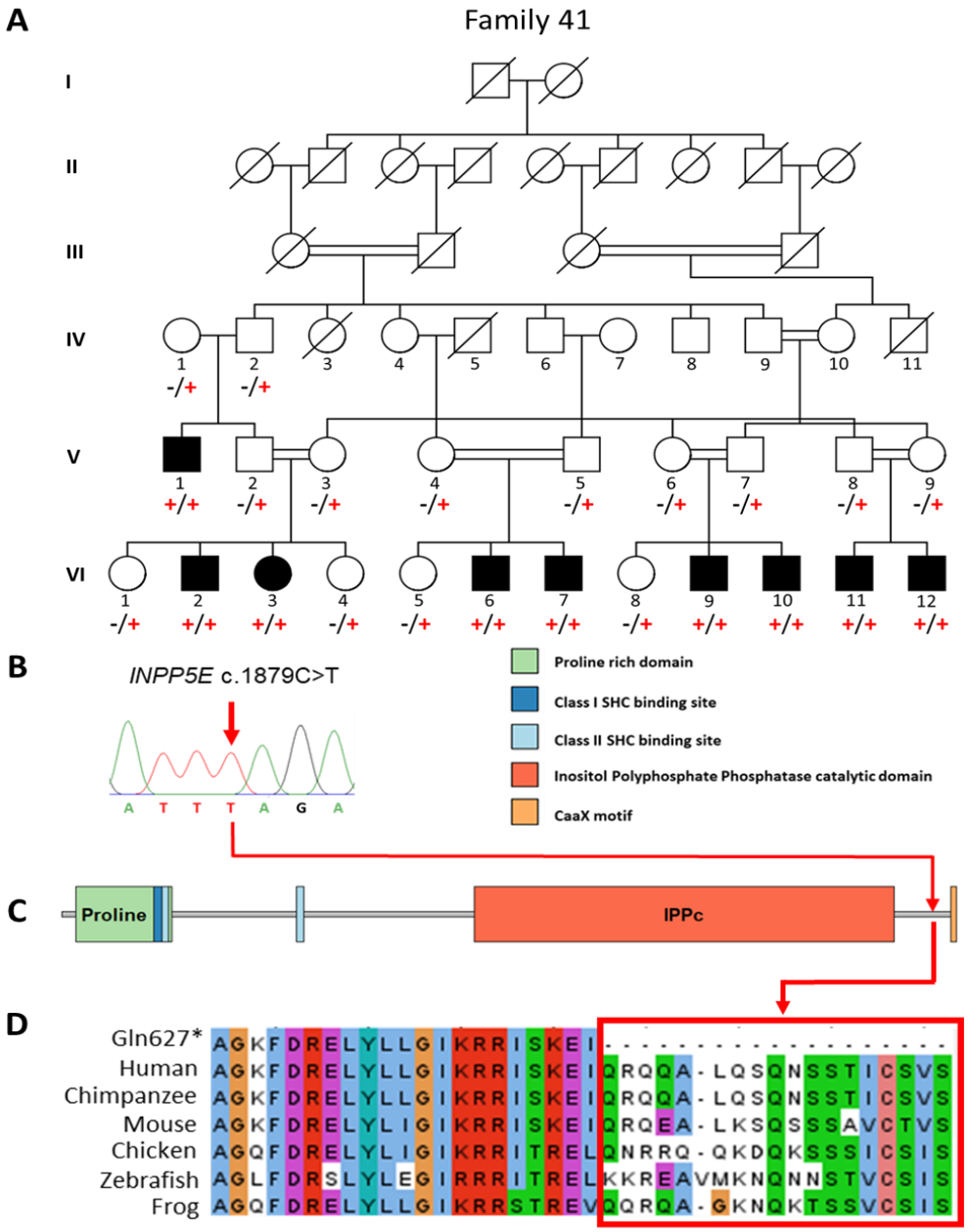


Figure 4.2 Family 41 pedigree showing *INPP5E* c.1879C>T genotype data

(A) Pedigree diagram of family 41 showing segregation of the INPP5E *c.1879C>T; p.(Gln627*) variant, co-segregating appropriately for an autosomal recessive condition. Genotypes are shown beneath generations IV to VI (+, c.1879C>T; -, wild type).*
(B) Sequence chromatogram of INPP5E *c.1879C>T in a homozygous affected individual.*
(C) Schematic showing domain architecture of INPP5E [adapted from Bielas et al *(250)]* and location of the *c.1879C>T; p.(Gln627*) variant. IPPc; Inositol Polyphosphate Phosphatase catalytic domain.*
(D) Conservation analysis: multiple species alignments of the partial amino acid sequences of INPP5E in a variety of vertebrate species. The p.(Gln627) variant is predicted to result in a truncated protein with deletion of the 18 C-terminal amino acids, including the highly conserved CaaX motif, but with preservation of the catalytic domain.*

Inpp5e knockout mice exhibit embryonic or perinatal lethality, and exhibit features of ciliopathy syndromes, including anophthalmia, polydactyly, polycystic kidneys, skeletal defects and cerebral developmental defects such as anencephaly and exencephaly (251). Post-natal *Inpp5e* knockout in 4-week-old mice results in further ciliopathy features including obesity, retinal degeneration and cystic kidneys (251). Additionally, ciliogenesis defects and abnormal cilia morphology have been noted in *Inpp5e* knockout mice embryos (251) as well as in *inpp5e* knockdown or knockout zebrafish embryos (256), supporting the importance of *INPP5E* in ciliogenesis and cilia maintenance, and linking dysfunctional ciliary phosphoinositol metabolism to the development of ciliopathy.

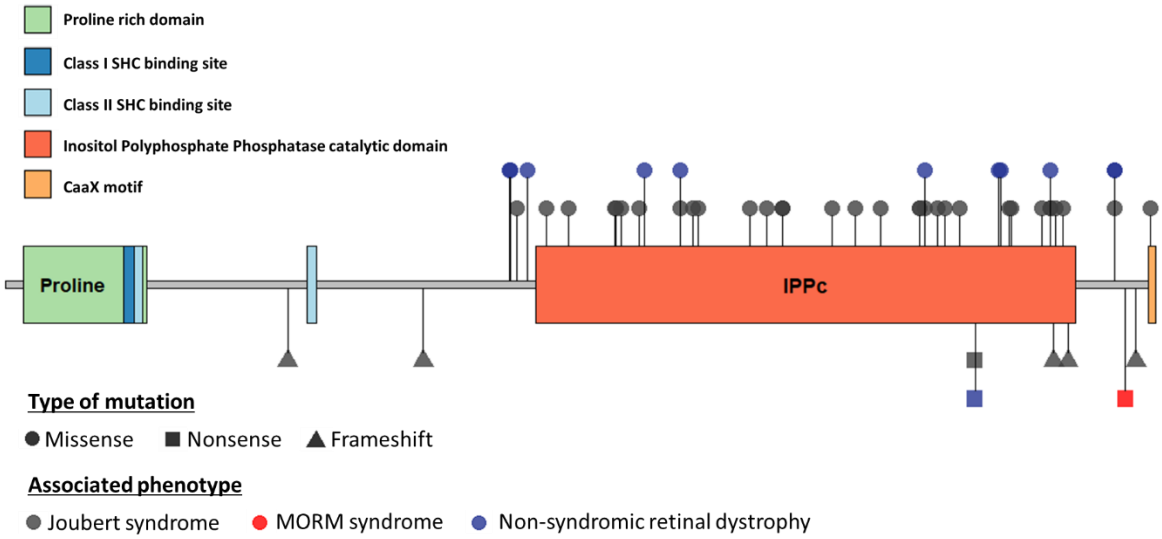


Figure 4.3 Localisation of *INPP5E* disease-associated variants

Schematic showing domain architecture of INPP5E *protein [adapted from Bielas* et al *(250)] with localisation of all reported disease-associated variants to date.*

Autosomal recessive pathogenic variants in *INPP5E* have been associated with two different ciliopathy disorders, JBTS and MORM syndrome (Table 4.2). Pathogenic variants in *INPP5E* are most commonly associated with JBTS, and these largely missense variants are clustered adjacent to or within the 5-phosphatase catalytic domain (Figure 4.3).

To date, there has only been a single variant associated with MORM syndrome, [*INPP5E* c.1879C>T; p.(Gln627*)], reported in only a single extended Pakistani family

with 14 affected individuals (251) (Table 4.2). This same variant was also identified in family 41 investigated in this study, who originate from the same province as the original reported family, and thus likely represents a regional founder variant in the Punjab region in Pakistan. Disease-associated *INPP5E* variants have also been reported in diverse populations of different ethnicities and geographical locations globally (Figure 4.4) (Table 4.2). In particular, two recurrent ('hot spot') gene variants have been described; the *INPP5E* p.(Arg435Gln) variant has been reported in six families from Turkey, Saudi Arabia, USA and Japan (250, 253, 257-259), whilst the p.(Arg621Gln) variant has been reported in six families from Algeria, Brazil, USA, Japan and Israel (253, 258, 260-263). Of interest, other variants disrupting the same amino acid residue as these recurrent *INPP5E* variants have also been described in affected individuals with INPP5E-associated disorders (264-266), suggesting the arginine at amino acid positions 435 and 621 may represent clinically significant residues.

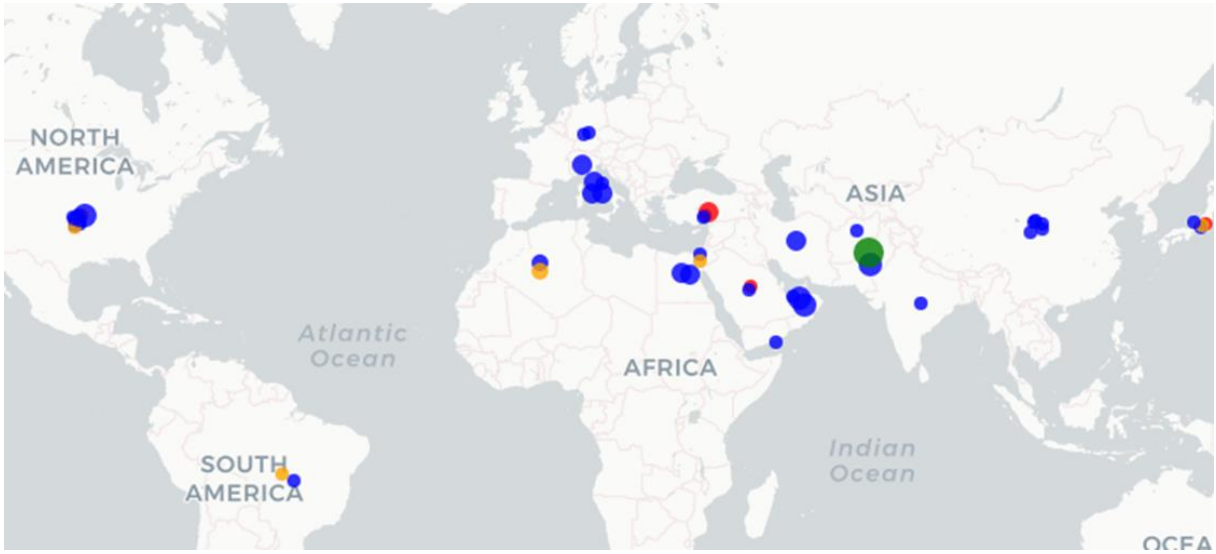


Figure 4.4 Geographical distribution of *INPP5E* disease-associated variants

Size of circles relates to the number of individuals identified with the same variant in the same country. The INPP5E *p.(Gln627*) MORM syndrome-associated variant is indicated by the green circle. Of note are the* recurrent INPP5E *p.(Arg435Gln) and p.(Arg621Gln) variants, highlighted by the red and orange circles respectively.*

The affected individuals in the original MORM syndrome family reported by Hampshire *et al* and Jacoby *et al* (249, 251) were described to have a constellation of clinical features that were suggestive of a ciliopathy disorder, but the phenotype was thought to be unique, and named MORM syndrome. Although patients said to have MORM

syndrome show significant clinical overlap with other ciliopathies, the authors distinguished MORM syndrome from JBTS and BBS on the basis of absence of polydactyly and the nature of visual impairment in MORM syndrome patients compared to patients with these other ciliopathy syndromes. In MORM syndrome, visual impairment due to retinal dystrophy presents early, but is non-progressive, whereas BBS tends to be associated with later onset retinal dystrophy and progressive visual impairment (249). Other distinguishing features of MORM syndrome include an absence of characteristic BBS facial features, and the presence of a small penis without testicular anomalies (249). The clinical features observed on examination of the nine affected individuals in this study were consistent with those in the previously reported family, consolidating intellectual disability (mild to moderate and non-progressive), obesity, visual impairment, aggressive and hyperactive behaviour, hypotonia and micropenis in males as core phenotypical features of MORM syndrome (Table 4.1).

The MORM syndrome-associated *INPP5E* nonsense variant is located within the last exon of the *INPP5E* gene, and is therefore predicted to escape nonsense-mediated decay, generating a prematurely truncated INPP5E protein lacking the terminal 18 amino acids with loss of the C-terminal CaaX domain (251). The CaaX motif (where C is a cysteine reside, aa are two small generally aliphatic residues, and X is any amino acid residue, contributing to its specificity) (267) acts as a signal for post-translational modification by farnesylation, leading to membrane targeting and attachment of the protein (268). Functional studies have demonstrated diminished capacity of the MORM syndrome-associated mutant truncated protein to stabilise the primary cilium, most likely due to a defect in ciliary localisation and protein interaction (251). Of particular note, this mutant protein still retains its catalytic domain with no impairment of enzymatic activity (251), in contrast to JBTS-associated missense mutations, where the mutant protein appears to localise correctly to the cilia, but is associated with reduced enzymatic activity (250, 269). It is tempting to postulate that this represents the molecular basis for the genotype-phenotype correlation associated with JBTS- and MORM-associated *INPP5E* mutations. Interestingly, apart from individuals with MORM syndrome, there has only been one other affected individual identified with biallelic truncating variants in *INPP5E* [homozygous for *INPP5E* c.1629C>G; p.(Tyr543*)] (253). This nonsense mutation was predicted to result in the production

of a truncated protein lacking the final part of the catalytic domain as well as the CaaX motif (although functional studies were not performed), similar to the MORM syndrome-associated variant, yet the affected individual had a clinical diagnosis of JBTS with the classic “molar tooth” sign on neuroimaging pathognomic of JBTS. In addition, De Goede *et al* reported an extended Pakistani family with 6 affected individuals all homozygous for an *INPP5E* p.(Tyr588Cys) missense variant displaying considerable phenotypic variability, with one individual demonstrating the classic “molar tooth” sign on neuroimaging, suggestive of a JBTS diagnosis, whilst another individual in the same family had a micropenis and clinical other features more consistent with MORM syndrome (269). It is important to note that it has not been practical to undertake neuroimaging in this study family, nor was it undertaken in the original family reported by Hampshire *et al* and Jacoby *et al* (249, 251), and so it remains unclear whether the cerebellum and brainstem are normal in the affected individuals with MORM syndrome, and hence the true extent to which MORM syndrome can be distinguished from JBTS.

Biallelic *INPP5E* variants have also been associated with isolated retinal dystrophies, including Leber congenital amaurosis, retinitis pigmentosa, macular and cone-rod dystrophy (260, 261, 266, 270-272) (Table 4.2). Interestingly, in at least four affected individuals, systemic features were specifically excluded, suggesting that some individuals with disease-causing *INPP5E* variants never develop the extraocular manifestations typically associated with JBTS or MORM syndrome (260, 266, 270, 271, 273). None of the individuals with *INPP5E*-associated isolated retinal dystrophy carry a combination of two loss-of-function variants, suggesting that complete loss of INPP5E function will always result in a systemic phenotype (Table 4.2). Apart from this observation, there does not seem to be a consistent correlation between an isolated retinal dystrophy or syndromic ciliopathy phenotype, and the nature of the variation (missense, nonsense or frameshift variants) (Figure 4.3). In fact, certain *INPP5E* variants p.(Arg515Trp), p.(Arg621Gln) and p.(Tyr543*) have been associated with both JBTS as well as isolated retinal dystrophy (250, 253, 258, 260-263, 270, 274) (Table 4.2).

Overall, these reports suggest that *INPP5E* variants may be associated with a wide range of ciliopathy phenotypes. Due to the phenotypic variability associated with

ciliopathies, it is often difficult to make diagnostic distinctions on the basis of clinical assessment alone, particularly in developing nations due to often-limited resources and social constraints. Identification of this *INPP5E* likely founder mutation will now enable targeted genetic testing for this variant in individuals in Northern Pakistan with an overlapping clinical presentation, permitting a rapid and cost-effective means of achieving a molecular diagnosis.

Table 4.2 Summary of all reported disease-associated *INPP5E* variants

Type of variant	Nucleotide variant	Protein variant	Associated Phenotype	Number of reported families (individuals)	Country of origin	Reference	ClinVar (Accession)
Missense	c.844G>A	p.(Gly282Arg)	Macular dystrophy	1 (1)	Germany	Birtel *et al* (270)	VUS (VCV000967675)
	c.848C>T	p.(Ala283Val)	LCA	1 (1)	China	Wang *et al* (266)	Not present
	c.856G>A	p.(Gly286Arg)	JBTS	1 (1)	Afghanistan	Travaglini *et al* (253)	Not present
	c.874C>G	p.(Arg292Gly)	RP	1 (1)	USA	Stone *et al* (261)	Likely pathogenic (VCV000857108)
	c.907G>A	p.(Val303Met)	JBTS	3 (3)	Italy (2); Uncertain (1)	Travaglini *et al* (253); Toma *et al* (275); Bachmann-Gagescu *et al* (276)	Pathogenic/ VUS (VCV000217653)
	c.944C>T	p.(Pro315Leu)	JBTS	2 (2)	USA (1); Uncertain (1)	Stone *et al* (261); Bachmann-Gagescu *et al* (276)	Pathogenic/ VUS (VCV000217662)
	c.1021G>A	p.(Gly341Ser)	JBTS	2 (3)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217661)
	c.1024T>C	p.(Cys342Arg)	JBTS	1 (1)	China	Chen *et al* (277)	Not present
	c.1035G>C	p.(Arg345Ser)	JBTS	1 (1)	Israel	Travaglini *et al* (253)	Not present
	c.1064C>T	p.(Thr355Met)	JBTS	2 (2)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217655)
	c.1073C>T	p.(Pro358Leu)	RP	1 (1)	China	Xu *et al* (271)	Not present
	c.1132C>T	p.(Arg378Cys)	JBTS	2 (3)	Italy	Bielas *et al* (250)	Pathogenic/ likely pathogenic (VCV000000400)
	c.1133G>A	p.(Arg378His)	LCA	1 (1)	Brazil	Porto *et al* (260)	Likely pathogenic (VCV000866268)
	c.1154G>A	p.(Cys385Tyr)	JBTS	1 (1)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217659)

c.1162G>T	p.(Val388Leu)	JBTS	1 (2)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217663)
c.1249T>C	p.(Ser417Pro)	JBTS	1 (1)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217664)
c.1277C>A	p.(Thr426Asn)	JBTS	1 (3)	Switzerland	Poretti *et al* (278)	Not present
c.1303C>G	p.(Arg435Gly)	JBTS	1 (1)	Turkey	Sönmez *et al* (264)	Not present
c.1303C>T	p.(Arg435Trp)	JBTS	1 (1)	India	Shetty et al (265)	Likely pathogenic (VCV000375472)
c.1304G>A	p.(Arg435Gln)	JBTS	6 (7)	Turkey (2); Saudi Arabia (1); USA (1), Japan (1); Uncertain (1)	Bielas *et al* (250); Travaglini *et al* (253); Alfares (257); Fleming *et al* (258); Suzuki *et al* (259); Bachmann-Gagescu *et al* (276);	Pathogenic (VCV000000399)
c.1388C>T	p.(Ala463Val)	JBTS	1 (1)	Saudi Arabia	Alfares (257)	Likely pathogenic (VCV000800892)
c.1426G>A	p.(Gly476Arg)	JBTS	1 (1)	Japan	Suzuki *et al* (259)	Not present
c.1468G>T	p.(Asp490Tyr)	JBTS	1 (1)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217665)
c.1534C>T	p.(Arg512Trp)	JBTS	2 (6)	Oman	Bielas *et al* (250)	Not present
c.1535G>A	p.(Arg512Gln)	JBTS	1 (1)	Yemen	Ben-Salem *et al* (274)	VUS (VCV000847803)
c.1543C>T	p.(Arg515Trp)	Cone-rod dystrophy	1 (1)	USA	Stone *et al* (261)	Pathogenic (VCV000000397)
		JBTS	3 (7)	Oman	Bielas *et al* (250); Ben-Salem *et al* (274)	
c.1565G>C	p.(Gly522Ala)	JBTS	3 (6)	USA	Hardee *et al* (252); Fleming *et al* (258); Summers *et al* (279)	Likely pathogenic (VCV000530891)
c.1577C>T	p.(Pro526Leu)	JBTS	1 (1)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic/ VUS (VCV000217660)
c.1600T>G	p.(Tyr534Asp)	JBTS	1 (2)	Algeria	Travaglini *et al* (253)	Not present
c.1668C>G	p.(Asp556Glu)	LCA	1 (1)	China	Wang *et al* (266)	Not present

	c.1669C>T	p.(Arg557Cys)	RP	1 (1)	China	Xu *et al* (271)	Likely pathogenic (VCV000426905)
	c.1684A>G	p.(Ser562Gly)	JBTS	1 (2)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217654)
	c.1688G>A	p.(Arg563His)	JBTS	2 (5)	Turkey (1); UAE (1)	Travaglini *et al* (253); Bielas *et al* (250)	Pathogenic (VCV000000398)
	c.1738A>G	p.(Lys580Glu)	JBTS	1 (3)	Egypt	Bielas *et al* (250)	Not present
	c.1753C>T	p.(Arg585Cys)	RP	1 (1)	USA	Stone *et al* (261)	Not present
			JBTS	2 (3)	Italy	Travaglini *et al* (253); Toma *et al* (275)	Not present
	c.1754G>A	p.(Arg585His)	JBTS	1 (1)	UAE	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217657)
	c.1763A>G	p.(Tyr588Cys)	JBTS	1 (6)	Pakistan	de Goede *et al* (269)	VUS (VCV000635031)
	c.1774C>G	p.(Arg592Gly)	JBTS	1 (1)	USA	Stone *et al* (261)	Not present
	c.1861C>T	p.(Arg621Trp)	LCA	1 (1)	Uncertain	Wang *et al* (273)	Likely pathogenic/ VUS (VCV000426904)
	c.1862G>A	p.(Arg621Gln)	LCA	2 (2)	Brazil (1); Israel (1)	Porto *et al* (260); Sharon *et al* (263)	Likely pathogenic (VCV000812336)
			Cone-rod dystrophy	1 (1)	USA	Stone *et al* (261)	
			JBTS	3 (4)	Algeria (1); Japan (1); USA (1)	Travaglini *et al* (253); Tsurusaki *et al* (262); Fleming *et al* (258)	
	c.1921T>C	p.(Cys641Arg)	JBTS	2 (4)	Egypt	Travaglini *et al* (253)	Not present
Nonsense	c.1629C>A	p.(Tyr543*)	Macular dystrophy	1 (1)	Germany	Birtel *et al* (270)	Not present
	c.1629C>G	p.(Tyr543*)	JBTS	1 (1)	Italy	Travaglini *et al* (253)	Not present
	c.1879C>T	p.(Gln627*)	MORM syndrome	2 (23)	Pakistan	Jacoby *et al* (251); Khan *et al* (61) – this study	Pathogenic (VCV000000396)

Frameshift	c.473delG	p.(Gly158Valfs*40)	JBTS	1 (1)	USA	Fleming *et al* (258)	Likely pathogenic (VCV000451128)
	c.1760delT	p.(Val587Glyfs*7)	JBTS	1 (2)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217656)
	c.1784_1787delTGAG	p.(Val595Glyfs*21)	JBTS	1 (1)	USA	Fleming *et al* (258)	Not present
	c.1897_1898delCA	p.(Gln633Glufs*64)	JBTS	1 (1)	Uncertain	Bachmann-Gagescu *et al* (276)	Pathogenic (VCV000217658)
	c.700dupC	p.(Leu234Profs*56)	JBTS	1 (1)	Japan	Tsurusaki *et al* (262)	Not present

Abbreviations: JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; RP, retinitis pigmentosa; UAE; United Arab Emirates; USA, United States of America; VUS, variant of uncertain significance. Ocular disorders are highlighted in blue, and the MORM syndrome is highlighted in yellow.

4.2.5 A novel *BBS5* variant associated with BBS in a Pakistani family

An additional multigenerational Pakistani family (family 42) residing in a remote village in the KPK province of Pakistan was also investigated. The two affected individuals in this family were also described as having clinical features suggestive of a ciliopathy disorder including moderate developmental delay/intellectual disability, visual impairment, truncal obesity, postaxial polydactyly, renal anomalies and hypogonadism, although this was phenotypically distinct from that of family 41 described above. WES identified the likely cause of disease as a novel homozygous frameshift mutation (NM_152384.2: c.196delA; p.(Arg66Glufs*12)) in *BBS5*. To date, there have only been two *BBS5* variants associated with BBS in Pakistani families, and this finding significantly expands on the contribution of *BBS5* mutations to BBS in Pakistan. A full description of the clinical and genomic investigations performed, as well as a comprehensive literature review of all published genetic causes of BBS in Pakistani families, can be found in Appendix B.

4.3 Phenotypic heterogeneity in an extended Amish family with BBS associated with homozygosity for the common *BBS1* p.(Met390Arg) variant

4.3.1 BBS

BBS is a rare, genetically heterogeneous, pleiotropic disorder, first described in the early 1920s by French paediatrician Georges Louis Bardet and Hungarian pathologist Arthur Biedl (235). BBS has an estimated prevalence of 1:125,000 - 160,000 in Northern Europe (280, 281), but is significantly more common in certain isolated communities such as the mixed Arab populations of Kuwait (1:36,000) (282) and in the Newfoundland population (1:18,000) (283). Primary features of BBS include retinal dystrophy, intellectual disability, postaxial polydactyly, obesity, renal dysfunction and hypogonadism (284). Retinal degeneration is the most highly penetrant feature of this condition (280, 285); in fact, after Usher syndrome, BBS is the second most common form of syndromic retinal degeneration (261). Other variable features of BBS include dental anomalies, cardiovascular defects, hearing loss, speech impairment, diabetes mellitus and other limb defects (284). To date, pathogenic variants in 21 genes have been associated with BBS in autosomal recessive or triallelic forms (286); some BBS genes have also been implicated in other related clinical presentations including syndromic ciliopathies as well as non-syndromic retinal dystrophies (Table 4.3).

Functionally, the majority of these genes either code for components of the core BBSome complex (*BBS1*, *BBS2*, *BBS4*, *BBS5*, *BBS7*, *TTC8*, *BBS9*, *BBIP1*), an octomeric protein complex involved in vesicular trafficking to the ciliary membrane (287), or for components of the chaperonin complex (*MKKS*, *BBS10*, *BBS12*), which plays an essential role in the assembly, stabilisation and regulation of the BBSome complex (288). Other BBS-associated genes code for proteins responsible for BBSome trafficking and localisation (*ARL6*, *CEP290* and *LZTFL1*) (289-291). The functions of some BBS genes, such as *C8orf37*, and their role in ciliary function, have not yet been fully characterised (292).

Table 4.3 Summary of genes associated with BBS

BBS type	BBS gene	Allelic disorders	Functions
BBS1	*BBS1*	-	Component of BBSome complex
BBS2	*BBS2*	Retinitis pigmentosa	Component of BBSome complex
BBS3	*ARL6*	Retinitis pigmentosa	GTPase; facilitates vesicular and interciliary trafficking
BBS4	*BBS4*	-	Component of BBSome complex
BBS5	*BBS5*	-	Component of BBSome complex
BBS6	*MKKS*	McKusick-Kaufman syndrome	Component of chaperonin complex
BBS7	*BBS7*	-	Component of BBSome complex
BBS8	*TTC8*	Retinitis pigmentosa	Component of BBSome complex
BBS9	*BBS9*	-	Component of BBSome complex
BBS10	*BBS10*	-	Component of chaperonin complex
BBS11	*TRIM32*	Limb-girdle muscular dystrophy	E3 Ubiquitin ligase activity
BBS12	*BBS12*	-	Component of chaperonin complex
BBS13	*MKS1*	Joubert syndrome; Meckel syndrome	Centriole migration
BBS14	*CEP290*	Joubert syndrome; Meckel syndrome; Senior-Loken syndrome; Leber congenital amaurosis	Interacts with RPGR
BBS15	*WDPCP*	Congenital heart defects, hamartomas of tongue, and polysyndactyly syndrome	Localisation of septins and ciliogenesis
BBS16	*SDCCAG8*	Senior-Loken syndrome	Regulates cell polarity, interacts with OFD1
BBS17	*LZTFL1*	-	Negatively regulates BBSome trafficking
BBS18	*BBIP1*	-	Component of BBSome complex
BBS19	*IFT27*	-	Intraflagellar transport
BBS20	*IFT74*	-	Ciliogenesis and length control
BBS21	*C8orf37*	Retinitis pigmentosa; Cone-rod dystrophy	Unknown

Adapted from (235, 286, 293).

To date, pathogenic alleles associated with overlapping ciliopathy phenotypes have been identified in only two genes in the Amish community; the *BBS1* (GRCh38) chr11:g.66526181T>G; NM_024649.4:c.1169T>G; p.(Met390Arg) variant, responsible for the widespread occurrence of BBS in the South-eastern Pennsylvania Amish community (294), and the *MKKS* variants chr20:g.10413265G>A; NM_170784.2:c.250C>T; p.(His84Tyr) and chr20:g.10412791C>A; c.724G>T; p.(Ala242Ser), which segregate in *cis* on a haplotype that is common in the Old Order Amish community, associated with the phenotypically distinct McKusick-Kaufman syndrome (295).

The present study entails clinical and genomic investigations in a large Amish family with multiple affected individuals displaying variable phenotypic features highly of BBS, in order to aid diagnosis and clinical management.

4.3.2 Materials and methods

Affected individuals and unaffected family members from an extended Amish family (family 43) were recruited to this study with informed consent. Blood and buccal sample collection and DNA extraction was performed as previously described (see section 2.3.2). WES was undertaken using DNA from a single affected individual in the family (Individual II:1) at BGI Hong Kong, as described in section 2.3.5. Bioinformatic analysis with additional virtual gene panel analysis and filtering to retain heterozygous variants compatible with triallelism was performed as described in section 4.2.2. Primer design, PCR and dideoxy sequencing (Appendix Table D2) was also performed as previously described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals.

4.3.3 Results: clinical and genetic findings

A large extended Amish family (family 43) with four affected individuals (II:1, II:2, II:5 and II:10) residing in Wisconsin (USA) was investigated. Clinical features in affected individuals included retinitis pigmentosa, polydactyly, obesity, frontal balding and dysmorphic facies.

To identify the causative mutation, targeted dideoxy genotyping was first performed in a single affected individual (II:1) for the pathogenic *BBS1* and *MKKS* variants previously identified in the Amish community [sequencing was only performed for the *MKKS* p.(His84Tyr) variant as a proxy for the p.(His84Tyr)/(Ala242Ser) double-variant haplotype known to be present in the Amish population] (294, 295). Genotyping results indicated that the affected individual II:1 was homozygous for the *BBS1* p.(Met390Arg) variant and wild type for the *MKKS* p.(His84Tyr) variant. The *BBS1* p.(Met390Arg) variant is present in gnomAD (v2.1.1 and v3.1.1), with a MAF of 0.00157 and 0.00205 respectively, higher in the non-Finnish European population (MAF 0.002773 - 0.002895) and in the Amish population (MAF 0.0307), and also present in a single individual in heterozygous form in a control exome dataset of 219 unrelated Amish individuals, although there are no homozygous individuals identified in either population database. The *BBS1* p.(Met390Arg) variant alters a highly conserved amino acid residue, and is predicted by multiple *in silico* tools (SIFT, Polyphen-2 and PROVEAN) to have a deleterious effect on protein function. This variant is reported as pathogenic or likely pathogenic in ClinVar (accession VCV000012143), and has been reported in homozygous and compound heterozygous form in multiple individuals with BBS (296, 297). Segregation analysis within the family confirmed that all affected individuals were homozygous for the *BBS1* p.(Met390Arg) variant; however, homozygosity for the *BBS1* p.(Met390Arg) variant was also identified in a single apparently unaffected individual (II:7) (Figure 4.5).

WES was subsequently performed in a single affected individual (II:1) in family 43 to identify additional variants that could be contributing to phenotypic penetrance under the digenic triallelic inheritance model of inheritance using the “rare multisystem ciliopathy disorders v1.84” PanelApp virtual gene panel. This confirmed homozygosity for the *BBS1* p.(Met390Arg) variant, but did not identify any further homozygous or compound heterozygous candidate gene variants compatible with the phenotype. Two additional candidate heterozygous variants of potential relevance were identified, a *BBS9* missense variant (GRCh38) chr7:g.33344585C>T; NM_198428.2:c.1280C>T; p.(Ala427Val) and a *CEP290* missense variant (GRCh38) chr12:g.88090784G>T; NM_025114.3:c.3517C>A; p.(Gln1173Lys), although due to time and Covid-19 laboratory constraints, segregation studies for both were not undertaken in the wider family.

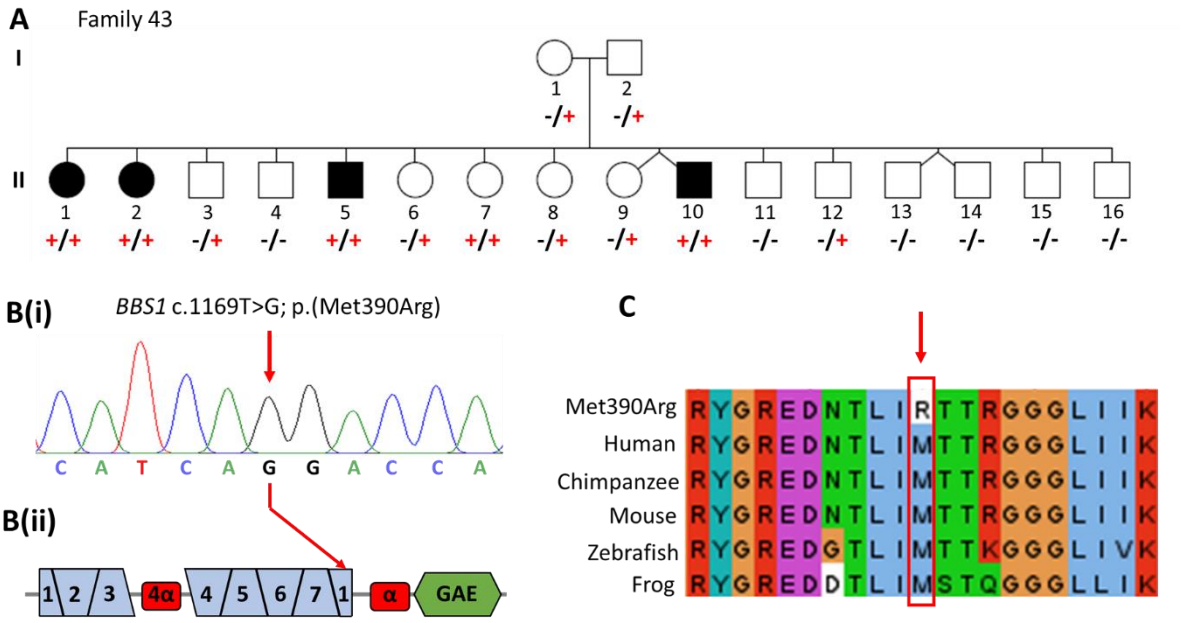


Figure 4.5 Family 43 pedigree showing *BBS1* p.(Met390Arg) genotype data

(A) Pedigree diagram of family 43 showing segregation of the BBS1 *c.1169C>T; p.(Met390Arg) variant, co-segregating appropriately for an autosomal recessive condition (apart from the apparently unaffected individual II:7). Genotypes are shown beneath generations I and II (+, c.1169C>T; -, wild type).*
(B)(i) Sequence chromatogram of BBS1 *c.1169C>T in a homozygous affected individual (ii) Schematic showing domain architecture of BBS1 [adapted from Chou* et al (298)*] and location of the c.1169T>G; p.(Met390Arg) variant. This variant is predicted to lie within the 7-bladed β-propeller domain, depicted in blue. The BBS1 protein is also predicted to contain a single α-helix as well as a 4α-helix bundle, depicted in red, and a γ-adaptin ear (GAE) domain, depicted in green.*
(C) Conservation analysis: multiple species alignments of partial amino acid sequences of BBS1 showing conservation of the Met390 amino acid residue across species.

The *BBS9* p.(Ala427Val) variant is present in homozygous form in 6 individuals in gnomAD (v2.1.1 and v3.1.1), with a MAF of 0.005462 - 0.005745, whilst the *CEP290* p.(Gln1173Lys) variant is absent in homozygous form in both gnomAD v2.1.1 and v3.1.1, but present in 19 and 13 individuals in heterozygous form respectively, with a MAF of 0.00008544 - 0.00009162. Neither variant has been associated with BBS in either autosomal recessive or triallelic forms. *In silico* predictions for both variants are inconsistent with regards to variant pathogenicity. ClinVar interpretations lean towards benign interpretation for *BBS9* p.(Ala427Val) variant (accession VCV000194138), whilst the *CEP290* p.(Gln1173Lys) variant is absent in ClinVar. Overall, the clinical significance of these two heterozygous variants remains unclear.

4.3.4 Discussion

BBS has traditionally been considered an autosomal recessive condition, established by segregation analysis of large pedigrees and consanguineous families. In some families however, more complex inheritance patterns have been proposed, which possibly account for some of the inter- and intra-familial phenotypic variability typically associated with BBS (299-301).

In the early 2000s, it was observed that a number of affected individuals who carried biallelic pathogenic variants in one of the BBS genes also carried an additional heterozygous variant in another BBS gene, and a “triallelic inheritance” model was proposed, suggesting that in some instances a third allele was either necessary for the disease to manifest (254, 302, 303), or instead exerted an epistatic effect on the BBS phenotype by modifying the age of onset and/or disease severity (304, 305). A number of BBS-associated genes have been reported to contribute to this complex digenic triallelic inheritance pattern, including *BBS1*, *BBS2*, *ARL6*, *BBS4*, *MKKS*, *BBS7*, *BBS10*, *BBS12* and *MKS1* (254, 303, 306-311). Additionally, variants in other ciliary genes, including *CCDC28B* (312), *ALMS1* (306), *NPHP4* (313), *TMEM67* (308), *RPGRIP1L* (314), *TTC21B* (315) and *NPHP1* (316) may also exert epistatic effects on BBS gene mutations and modify severity of the BBS phenotype. Evidence from a homozygous hypomorphic *Cep290* mutant mouse model, where the additional loss of *Bbs4* alleles results in increased obesity and accelerated retinal degeneration compared with mice without *Bbs4* mutations, supports an interaction between BBSome components and ciliary proteins in modifying BBS phenotypic severity (317). An increased mutational load leading to dysfunction of more than one component in this ciliary protein trafficking pathway could result in an additive impairment of ciliary function, providing a plausible mechanistic basis for the variable penetrance or phenotypic severity seen in cases of BBS with triallelic inheritance (318).

To date, however, the triallelic inheritance hypothesis in BBS remains controversial. There are a few reports of non-penetrant individuals who carry clearly deleterious biallelic variants in a single BBS gene, and yet are apparently unaffected (254, 303, 319, 320). Most of these reports involve the same *BBS1* p.(Met390Arg) variant

identified in this study, and the presence of a single apparently unaffected homozygous individual in this study (individual II:7) is therefore not unprecedented. The *BBS1* p.(Met390Arg) variant is one of the most common pathogenic variants in the *BBS1* gene, and has been reported in homozygous and compound heterozygous state in multiple affected BBS individuals (296, 297), accounting for up to 80% of BBS1 cases in European populations (297). A *Bbs1* p.(Met390Arg) knock-in mouse model recapitulates features of the human BBS phenotype, including retinal degeneration, male infertility and obesity (321), clearly defining this as a pathogenic BBS variant (ClinVar accession VCV000012143.26). These findings have been used to support the triallelic hypothesis, suggesting that at least in some cases a third pathogenic allele is required for the disease to manifest. Indeed, Badano and coworkers initially reported two BBS families where both the affected individual as well as the unaffected father were found to be homozygous for the *BBS1* p.(Met390Arg) variant (303). In one of these families, the differential penetrance was subsequently explained by the epistatic contribution of a heterozygous *CCDC28B* NM_024296.5:c.330C>T variant identified in the affected individual but not in the unaffected father (312). *CCDC28B* codes for a pericentriolar protein that interacts and co-localises with BBS proteins including BBS1, BBS2, BBS4, BBS5, BBS6, BBS7, BBS8 (312). The *CCDC28B* c.330C>T variant is a synonymous variant p.(Phe100Phe) that enhances the use of a cryptic splice acceptor site, thus introducing a premature termination codon and reducing *CCDC28B* mRNA levels (312). This variant however was not identified in this study.

Functional studies in zebrafish indicate the *BBS1* p.(Met390Arg) variant may be a hypomorphic variant, as evidenced by partial phenotypic rescue following introduction of the variant into zebrafish embryos lacking *BBS1* expression (322). This could explain the milder disease phenotype associated with homozygosity for the *BBS1* p.(Met390Arg) variant, with a later age of disease onset (306, 310, 323) and a clinical presentation that may be limited to retinal degeneration with minor (polydactyly or obesity only) or even none of the systemic features commonly associated with BBS (319, 323-326). In fact, there have been reports of *BBS1* p.(Met390Arg) homozygous individuals in their fourth decade of life who were asymptomatic at the time of presentation, with genetic investigations only prompted by a routine eye examination or a relative with a retinal disease diagnosis (323). An alternative explanation for the apparent non-penetrance of the *BBS1* p.(Met390Arg) variant in individual II:7 in this

study (aged 12 at the time of examination) is that she may actually harbour a very mild subclinical phenotype; a further detailed ophthalmic examination would be useful to clarify this, although this was unfortunately not possible. It is still difficult to definitely ascertain if there are truly non-penetrant individuals with biallelic pathogenic BBS variants, particularly with regards to the hypomorphic *BBS1* p.(Met390Arg) variant.

There are also numerous studies that do not support a triallelic inheritance pattern modifying either penetrance or expressivity (296, 297, 327-331). Although BBS is a rare disorder in non-endogamous populations, the large number of BBS-associated genes suggests that the cumulative carrier frequency for at least one BBS mutation could be as high as 1 in 50 (331), and additional rare heterozygous BBS variants in affected individuals are possibly being detected by chance alone. Additionally, the third alleles reported in the literature are commonly missense variants, for which pathogenicity can sometimes be difficult to evaluate.

Adding further to the genetic complexity in BBS, there is experimental evidence that some heterozygous modifier BBS variants may influence disease phenotype via a dominant-negative effect, and this may even be the primary mechanism by which modifier alleles exacerbate disease phenotype (322). This predicts that carriers of dominant negative alleles may manifest subclinical BBS phenotypes. Consistent with this hypothesis, BBS carriers have been reported to be at increased risk of retinal dysfunction, obesity, hypertension, diabetes and renal disease (332-335), although this finding has also been refuted by others (336).

Given the biological and mechanistic overlap of many BBS genes in ciliary function, it is quite plausible that the phenotypic variability often observed in BBS could in part be due to the modifying effects of sequence variants in genes encoding other ciliary proteins, and in reality is likely to reflect a complex interaction between multiple genetic factors and environmental influences. It may well be that with continued improvements in the identification and functional validation of deep intronic, regulatory region and copy number variants, combined with detailed phenotypic characterisations, the impact of additional modifier gene variants in BBS can be further clarified (337, 338). This will ultimately allow more informative counselling of affected families with regards to prognosis and recurrence risk.

4.4 Delineating the expanding phenotype associated with *SCAPER* gene mutation

4.4.1 SCAPER syndrome

SCAPER (S phase cyclin A–associated protein residing in the ER), encoded by the *SCAPER* gene on chromosome 15q24.3, is a cyclin A/Cdk2 regulatory protein that interacts with the cyclin A/Cdk2 complex at multiple phases of the cell cycle, and is involved in cell cycle progression (339). A potential role for *SCAPER* in human disease was first suggested by Najmabadi *et al*, who identified a candidate homozygous frameshift *SCAPER* variant as the cause of non-syndromic intellectual disability in a small Iranian family (340). Since then, biallelic *SCAPER* variants have been identified in 24 individuals presenting with a novel syndromic disorder characterised by intellectual disability and retinitis pigmentosa in association with variable multisystem phenotypical presentations.

Here, clinical and genetic findings, including seven novel *SCAPER* variants, are described in six individuals in five families (families 44-48) of Amish, Caucasian and South Asian descent. Together with molecular data and comprehensive phenotypic assessments, this enabled a more detailed clinical comparison to be drawn between the patient cohort described [including previously published individual G001284; Patient 3 (family 45) in this study (85)] with the 24 individuals in whom *SCAPER* variants were recently defined (340-345), permitting a more precise definition of the clinical phenotype arising from pathogenic *SCAPER* variation.

4.4.2 Materials and methods

Samples were taken with informed consent for DNA extraction (see section 2.3.2). SNP genotyping was performed (Patients 1 and 2; family 44) as described in section 2.3.4. WES was undertaken as per section 2.3.5, with bioinformatics analysis, primer design, PCR and dideoxy sequencing (Appendix Table D2) performed as described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals. Patient 6 (family

48) underwent WES at GeneDx and was identified via GeneMatcher (346) as part of the Matchmaker Exchange Repositories (347).

A literature review was performed as described in section 2.4 to retrieve all reported *SCAPER* disease-associated variants. Findings are summarised in Tables 4.4 and 4.5.

4.4.3 Results: clinical and genetic findings

Tables 4.4 and 4.5 summarise the core phenotypical features of individuals in this study, aged between 18 months and 31 years (Patients 1 - 6; families 44 - 48), and compares these to the clinical features of all SCAPER syndrome patients described to date. Intellectual disability and developmental delay was present in all six affected individuals, and four patients also exhibited hyperactivity and attention deficit hyperactivity disorder (ADHD). Autism and dyspraxia were each noted in one individual. Neuroimaging performed in patients 1, 3, 5 and 6 revealed no abnormalities. Additional dysmorphic features noted in both Amish siblings (patients 1 and 2) included inverted nipples, brachydactyly, camptodactyly, proximally placed thumbs (Figure 4.6) and a characteristic facial appearance with frontal bossing and almond-shaped eyes; growth parameters were all normal. Patients 1, 3 - 6 all presented between the ages of 10 - 23 with a reduction in night vision and visual field deficits; Patient 2 described no visual symptoms at the time of presentation. Fundus examination in patients 3 - 6 revealed findings typical of retinitis pigmentosa including optic disc pallor, attenuated retinal vessels and intraretinal mid-peripheral bone-spicule pigmentation and loss of photoreceptor outer segments with retained central macular structure on OCT imaging (Figure 4.6; Table 4.5). Additional variable ocular features described in some patients with SCAPER syndrome include cataracts (in two individuals) and myopia and keratoconus in one individual each (Table 4.5).

Table 4.4 Comparison of clinical findings of all affected individuals with biallelic pathogenic *SCAPER* variants

	Genotype	Ethnicity	Gender	Age (yr)	Wt kg (SDS)	Ht cm (SDS)	OFC cm (SDS)	BMI (SDS)	Walked	Speech delay	ID	Behaviour issues	Abnormal neuro-imaging	RP	Brachy-dactyly	Other clinical findings
Najmabadi	p.(Tyr118fs*)/ p.(Tyr118fs*)	Iran	NA	NA	NA	NA	NA	NA	NA	NA	✓	NA	NA	NA	NA	NA
Tatour (A:II:1)	c.2023-2A>G/ c.2023-2A>G	Arab	F	24	NA	NA	NA	NA	Normal	NA	Mild (IQ 64)	ADHD	MRI: normal	✓	NA	Nil
Tatour (A:II:2)	c.2023-2A>G/ c.2023-2A>G	Arab	F	23	NA	NA	NA	NA	Normal	NA	Mild (IQ 56)	ADHD	NA	✓	NA	Nil
Tatour (B:II:1)	c.2973_2976del; p.(Ile991fs*)/ c.2973_2976del; p.(Ile991fs*)	Spanish	F	34	NA	NA	NA	NA	24 mo	NA	Mod	None reported	CT: normal	✓	NA	Alopecia Areata
Tatour (C:II:4)	c.1859_1861del; p.(Glu620del)/ c.3565G>A; p.(Ser1219Asn)	Spanish	M	15	NA	NA	NA	NA	Delayed	NA	✓	None reported	NA	✓	NA	NA
Jauregui	c.2023-2A>G/ c.2023-2A>G	Arab	M	11	NA	NA	NA	NA	NA	NA	No	No	NP	✓	NA	NA
Wormser (P1:V5)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	F	34	78 (+1.9)	145 (-3.1)	Not reduced	37.1 (+3.1)	NA	✓	Mod	NA	NP	✓	✓	Genu valgum/ genu varum
Wormser (P1:V6)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	M	28	78 (+0.7)	157 (-3.1)	Not reduced	31.6 (+2.3)	NA	✓	Mod	NA	NP	✓	✓	Genu valgum/ genu varum
Wormser (P1:V7)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	M	24	98 (+2.2)	163 (-2.2)	Not reduced	36.9 (+3.0)	NA	✓	Mod	NA	NP	✓	✓	Genu valgum/ genu varum
Wormser (P1:V8)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	M	17	92 (+2.2)	155 (-2.9)	Not reduced	38.3 (+3.3)	NA	✓	Mod	NA	NP	✓	✓	Genu valgum/ genu varum
Wormser (P2:III1)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	F	48	86.6 (+2.5)	146 (-3.0)	Not reduced	40.6 (+3.5)	NA	✓	Sev	NA	NP	✓	✓	Nil
Wormser (P2:III2)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	F	47	62 (+0.4)	149 (-2.5)	Not reduced	27.9 (+1.6)	NA	✓	Sev	NA	NP	✓	✓	Genu valgum/ genu varum
Wormser (P2:III7)	c.2806del; p.(Leu936*)/	Bedouin	F	29	57.8 (+0.1)	132 (-5.2)	Not reduced	33.2 (+2.8)	NA	✓	Sev	NA	NP	✓	✓	Nil

	c.2806del; p.(Leu936*)															
Wormser (P2:IV1)	c.2806del; p.(Leu936*)/ c.2806del; p.(Leu936*)	Bedouin	M	10	29.5 (-0.38)	129 (-1.5)	Not reduced	17.7 (+0.7)	NA	✓	Mod	ADHD	MRI: abnormal‡	suspected	✓	Genu valgum/ genu varum
Kahrizi (1:III:6)	Exon 15-16del	Pakistani	M	18	42 (<3rd cent)	158 (<3rd cent)	55 (13th cent)	16.8	21 mo	✓ 3 yrs	Mod (IQ 50)	Self-mutilation	MRI/ EEG: normal	✓	NA	Prominent nose, narrow chin, high forehead, 2nd/3rd toe syndactyly, funnel chest, hypotonia, polyneuropathy UL & LL, single episode febrile seizure 18 mo
Kahrizi (1:III:7)	Exon 15-16del	Pakis-tani	F	12	43.5 (44th cent)	150 (26th cent)	55 (76th cent)	19.1	21 mo	2 yrs	Mild (IQ 67)	No	MRI: mild asymmetry lateral ventricles	✓	NA	Prominent nose, narrow chin, high forehead, hepatomegaly polyneuropathy UL & LL
Kahrizi # (2:II:1)	c.1096C>T; p.(Arg366*)/ c.1096C>T; p.(Arg366*)	Baloch	F	34	NA	152 (-2.0)	52.5 (0)	NA	3 yrs	✓ 4 yrs	Sev (IQ 31)	No	NP	✓	NA	Prominent maxilla & micrognathia, generalised tonic clonic seizures in infancy
Kahrizi # (2:II:2)	c.1096C>T; p.(Arg366*)/ c.1096C>T; p.(Arg366*)	Baloch	M	32	NA	160 (-2.3)	51 (-2.8)	NA	3 yrs	✓ 4 yrs	Sev (IQ 30)	No	NP	✓	NA	Prominent maxilla & micrognathia, seizures in infancy
Kahrizi # (2:II:4)	c.1096C>T; p.(Arg366*)/ c.1096C>T; p.(Arg366*)	Baloch	M	26	NA	172 (+1.0)	55 (0)	NA	3 yrs	✓ 3 yrs	Sev (IQ 34)	No	NP	✓	NA	Prominent maxilla, generalised tonic clonic seizures in infancy
Kahrizi (3:IV:1)	c.1092dup; p.(Val365fs*)/ c.1092dup; p.(Val365fs*)	NA	M	32	NA	164 (-1.8)	57.5 (+2.0)	NA	2.5 yrs	2 yrs	Mod (IQ 40)	No	NP	✓	NA	Partial vitiligo on extremities
Kahrizi (3:IV:2)	c.1092dup; p.(Val365fs*)/ c.1092dup; p.(Val365fs*)	NA	F	12	40	145 (-1.0)	55 (+2.0)	21.4	18 mo	18 mo	Mod (IQ 45)	No	NP	✓	NA	Nil
Kahrizi (3:IV:3)	c.1092dup; p.(Val365fs*)/ c.1092dup; p.(Val365fs*)	NA	M	20	NA	163 (0)	55 (+2.0)	NA	12 mo	24 mo	Mod (IQ 40)	No	NP	✓	NA	Nil

Kahrizi (3:IV:5)	c.1092dup; p.(Val365fs*)/ c.1092dup; p.(Val365fs*)	NA	M	25	NA	164 (0)	52.5 (0)	NA	13 mo	24 mo	Mod (IQ 50)	No	NP	✓	NA	Nil
Kahrizi (4:VI:1)	c.1883T>G; p.(Phe628Cys)/ c.1883T>G; p.(Phe628Cys)	NA	F	7	15	112 (0)	49 (-2.0)	12	2.5 yrs	✓ 3 yrs	Mild (IQ 60)	No	NP	✓	NA	Upslanting palpebral fissures, epicanthal fold, small mouth, low set ears
Family 44 Patient 1	c.2236dup; p.(Ile746fs*)/ c.2236dup; p.(Ile746fs*)	Amish	M	13.7	68.9 (+1.9)	166.3 (+0.7)	56.4 (+0.4)	24.9 (+2.0)	24 mo	✓	Mod	Hyper-activity	MRI: normal	No	✓	Proximally placed thumbs, short 5th fingers, pes planus, frontal bossing, almond-shaped eyes, inverted nipples
Family 44 Patient 2	c.2236dup; p.(Ile746fs*)/ c.2236dup; p.(Ile746fs*)	Amish	F	1.5	8.6 (-2.2)	78.5 (-0.7)	47.0 (-0.9)	14.0 (-2.5)	22 mo	✓	Mild	Hyper-activity	NP	NA (age)	✓	Proximally placed thumbs, short 5th fingers, pes planus, frontal bossing, almond-shaped eyes, inverted nipples
Family 45, Patient 3*	c.2179C>T; p.(Arg727*)/ c.1116del; p.(Val373fs*)	South Asian	F	28	25th cent	3rd cent	NA	NA	11 mo	✓	Mod	ADHD Autism Self-harm	MRI: norm	✓	NA	Nil
Family 46, Patient 4	c.1495+1G>A/ c.3224del; p.(Pro1075fs*)	Caucasian	F	31	NA	NA	57 (95th cent)	NA	15 mo	✓	Mild	Dyspraxia	NP	✓	NA	Nil
Family 47, Patient 5	c.829C>T; p.(Arg277*)/ c.3707_3708del; p.(Ser1236fs*)	NA (USA)	F	17	NA	NA	NA	Obese	NA	NA	✓	NA	MRI: norm	✓	NA	Nil
Family 48, Patient 6	c.2377C>T; p.(Gln793*)/ c.2166-3C>G	NA (USA)	F	24	63.6 (+0.6)	162.6 (-0.2)	NA	24.0 (+0.6)	15-18	✓	Mild (IQ 50-60)	ADHD	MRI: norm	✓	NA	Mod eczema with sev skin picking behavior
Summary								8/16 obese		18/23	29/30	9/21		26/27	10/10	

Abbreviations: abnorm, abnormal; ADHD, attention-deficit hyperactivity disorder; cent, centile; CT, computerised tomography; F, female; Ht, height; ID, intellectual disability; IQ, Intelligence quotient (Wechsler Adult Intelligence Scale); LL, lower limb; M, male; mo, months; Mod, moderate; MRI, magnetic resonance imaging; NA, not available; NP, not performed; OFC, occipitofrontal circumference; RP, retinitis pigmentosa; SDS, standard deviation scores; Sev, severe; susp, suspected; UL, upper limb; vent, ventricles; Wt, weight; yrs, years. (✓) indicates presence of a feature in an affected subject.

[†]Refers to age of examination
[‡]Abnormal MRI findings include: mildly enlarged lateral ventricles and several loci of irregular signal in the brain parenchyma above the tentorium, in the posterior white matter and along the ependyma.

**Also patient G001284 (Carss* et al *2017 and Turro et al 2020 (85, 348))*
[#]Also family M8500314 (Hu et al *2018 (344))*

Height, weight, BMI and OFC Z-scores were calculated using a Microsoft Excel add-in to access growth references based on the LMS method (349) using a reference European population (350).
Adults with a BMI >25 are classified as overweight, those >30 are classified as obese.

Table 4.5 Ocular findings of all affected individuals with biallelic pathogenic *SCAPER* variants

	Presenting age (yrs)	Nyctalopia	↓ visual fields	VA logMAR (Snellen)	Strabismus	Cataract (morphology)	Optic disc pallor/ atrophy	Retinal vessel attenuation	Retinal pigmentary changes	CMO	Electrophysiology	Other ocular findings
Najmabadi	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Tatour (A:II:1)	12-13	✓	NA	Mod/Sev ↓	NA	✓	✓	✓	✓	✓	Undetectable rod and cone responses	Secondary glaucoma
Tatour (A:II:2)	12-13	✓	NA	Mod/Sev ↓	✓	No	✓	✓	✓	No	Undetectable rod and cone responses	Nil
Tatour (B:II:1)	28	✓	✓	Mod ↓	NA	✓ (PSC)	✓	✓	✓	✓	Extinguished rod and cone responses	Nystagmus High myopia
Tatour (C:II:4)	15	NA	✓	↓	NA	NA	✓	✓	✓	No	Abolished	Nil
Jauregui	9	✓	NA	RE: 0.3 (20/40) LE: 0.1 (20/25)	NA	NA	✓	✓	✓	No	Undetectable rod, subnormal cone responses	Nil
Wormser (P1:V5)	10	✓	NA	RE: PL LE: HM	✓	✓ (PSC)	✓	✓	✓	NA	Extinguished rod and cone responses	Nil
Wormser (P1:V6)	15	✓	NA	RE: 0.3 (20/40) LE: 0.3 (20/40)	✓	✓ (PSC, nuclear, punctate)	✓	✓	✓	NA	Extinguished rod and cone responses	Nil
Wormser (P1:V7)	13	✓	NA	RE: 1.0 (20/200) LE: 2.0 (20/400)	✓	✓ (PSC)	✓	✓	✓	NA	NA	Nil
Wormser (P1:V8)	7	✓	NA	RE: 0.24 (20/33) LE: 0.24 (20/33)	No	No	NA	NA	Suspected	NA	NA	Nil
Wormser (P2:III1)	20	✓	NA	RE: NPL LE: NPL	No	✓ (mild cortical)	✓	✓	✓	✓	NA	Nil
Wormser (P2:III2)	28	✓	NA	RE: HM 15cm LE: PL	✓	✓ (mild PSC)	✓	✓	✓	✓	NA	Nil
Wormser (P2:III7)	25	✓	NA	RE: 0.54 (20/70) LE: 0.54 (20/70)	No	No	No	No	✓	NA	NA	Nil
Wormser (P2:IV1)	NA	Unable to assess	NA	Fixes and follows objects	No	No	✓	No	✓	NA	NA	Nil
Kahrizi (1:III:6)	Adolescence	NA	✓	NA	No	No	NA	NA	✓	NA	NA	Myopia (RE -7.25D, LE -8.0D)
Kahrizi (1:III:7)	12	NA	NA	↓	No	No	NA	NA	✓	NA	NA	Myopia (RE -12.25D, LE -10.25D)

Kahrizi [#] (2:II:1)	NA	NA	NA	↓	No	No	NA	NA	✓	NA	NA	Nil
Kahrizi [#] (2:II:2)	NA	NA	NA	NA	✓	No	NA	NA	✓	NA	NA	Nil
Kahrizi [#] (2:II:4)	NA	NA	NA	NA	No	No	NA	NA	✓	NA	NA	Nil
Kahrizi (3:IV:1)	NA	✓	NA	NA	✓	No	NA	NA	✓	NA	NA	Nil
Kahrizi (3:IV:2)	NA	✓	NA	NA	✓	No	NA	NA	✓	NA	NA	Nil
Kahrizi (3:IV:3)	NA	✓	NA	NA	✓	No	NA	NA	✓	NA	NA	Nil
Kahrizi (3:IV:5)	NA	✓	NA	NA	✓	No	NA	NA	✓	NA	NA	Nil
Kahrizi (4:VI:1)	NA	✓	NA	NA		No	NA	NA	✓	NA	NA	Nil
Family 44 Patient 1	13	✓	No	RE: 0.48 (20/60) LE: 0.48 (20/60)	✓	No	No	No	No	No	NP	Bilateral meridonal amblyopia
Family 44 Patient 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Family 45, Patient 3*	14	✓	✓	RE: 0.3 (20/40) LE: 0.3 (20/40)	NA	No	✓	✓	✓	No	Undetectable rod responses, pERG subnormal	Keratoconus
Family 46, Patient 4	10	✓	NA	RE: 0.2 (20/30) LE: 0.3 (20/40)	NA	✓	✓	✓	✓	NA	Undetectable rod responses	Myopia
Family 47, Patient 5	NA	✓	✓	NA	NA	NA	✓	✓	✓	✓	Severe ↓	Nil
Family 48, Patient 6	23	Unable to definitively assess	✓	RE: 0.2 (20/30) LE: 0.3 (20/40)	✓	✓ (nuclear)	NA	✓	✓	✓	Undetectable rod, severely ↓ cone responses	Nil
Summary		20/20	5/6		12/21	9/24	14/16	14/17	26/27	6/11	11/11	

Abbreviations: CMO, cystoid macular oedema; HM, detection of hand motion; LE, left eye; Mod, moderate; NA, not available; NPL; no perception of light; pERG, pattern ERG; PL, perception of light; PSC, posterior subcapsular cataract; RE, right eye; Sev, severe; VA, visual acuity; ↓, reduced. (✓) indicates presence of a feature in an affected subject

**Also patient G001284 (Carss* et al *2017 and Turro* et al *2020 (85, 348)).*
[#]Also family M8500314 (Hu et al *2018 (344)).*

Patients 1 and 2 from family 44 are Ohio Amish siblings. Candidate variants identified through WES of DNA from patient 1 were cross-referenced with regions of autozygosity common to both affected siblings, identified through whole genome SNP genotyping. This identified only a single plausible candidate variant, located within the largest (18 Mb) shared region of autozygosity on chromosome 15 [rs1509805 - rs4243078; (GRCh38) chr15:g.60281446 - 78374545], a novel homozygous duplication variant in exon 18 of the *SCAPER* gene, predicted to result in a frameshift [(GRCh38) chr15:g.76705914dupT, NM_020843.2:c.2236dupA, p.(Ile746Asnfs*6); Figure 4.6]. Dideoxy sequencing confirmed the presence and co-segregation of this variant in both siblings. This variant was detected in heterozygous form in five unrelated individuals in a database of 116 regional Amish controls, corresponding to an estimated allele frequency of ~0.04, not uncommon for founder mutations within this population.

In patients 3 - 6, NGS identified compound heterozygous *SCAPER* variants summarised in Table 4.4. The *SCAPER* variants in each of these patients were confirmed to be biallelic by familial segregation analysis using dideoxy sequencing. *SCAPER* variants identified in Patients 1, 2, 4 - 6 are novel. The p.(Ile746Asnfs*6) and p.(Gln793*) *SCAPER* variants are present in gnomAD (v3.1.1) a single (non-Finnish) European individual, the p.(Arg727*) variant in a single South Asian individual (gnomAD v3.1.1), and the p.(Ser1236Tyrfs*28) in a single African/African-American individual (gnomAD v2.1.1 and v3.1.1), all in heterozygous form only; remaining SCAPER variants were absent from the gnomAD population database.

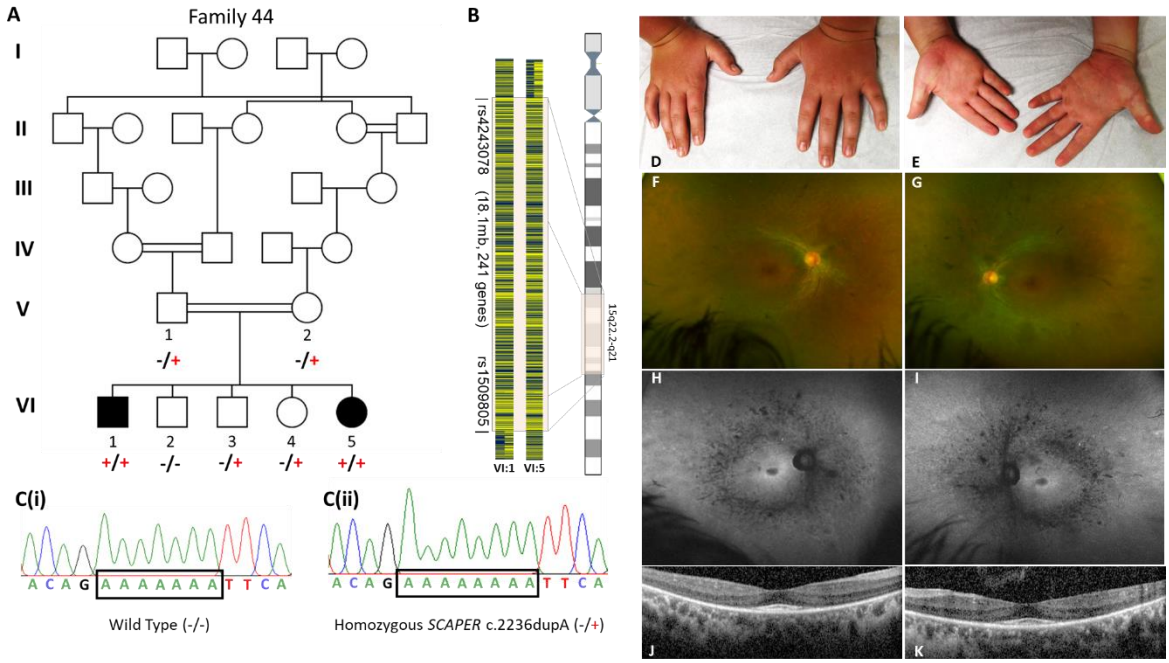


Figure 4.6 Amish (family 44) pedigree showing *SCAPER* c.2236dupA, p.(Ile746Asnfs*6) genotype data and selected clinical images of affected individuals

(A) Amish (family 44) pedigree (patient 1; VI:1 and patient 2; VI:5) showing segregation of the SCAPER *c.2236dupA; p.(Ile746Asnfs*6) variant, co-segregating appropriately for an autosomal recessive condition. Genotypes are shown beneath generations V and VI (+, c.2236dupA; -, wild type).*
(B) Pictorial representation of SNP genotypes across the ~18.1 Mb chromosome 15q21-22 region identified in the Amish family.
(C) Sequence chromatogram of SCAPER *c.2236dupA in (i) an unaffected wild type individual and (ii) a homozygous affected individual.*
(D-K) Clinical features of SCAPER syndrome patients. (D-E) Brachydactyly, camptodactyly and proximally placed thumbs in Patient 1. (F-K) Ocular imaging in Patient 3 illustrating features of retinitis pigmentosa. Fundus photograph (Optos California) of right (F) and left (G) eyes showing optic disc pallor, attenuated retinal vessels and mid-peripheral bone spicule pigmentation. Fundus autofluorescence (FAF) imaging of right (H) and left (I) eyes showing mid-peripheral hypoautofluorescence with a central ring of hyperautofluorescence demarcating the surviving outer retinal structures. Optical coherence tomography (Spectralis-OCT) of right (J) and left (K) eye central retina demonstrating loss of photoreceptor outer segments with retained central macular structure corresponding to FAF findings.

4.4.4 Discussion

Here clinical and genetic studies in six affected individuals (including additional new clinical details for Patient 3 (85)) were undertaken, which takes the total number of SCAPER syndrome patients described to date to 30. Although the extent for which clinical data is available for the previously reported SCAPER patients is variable, the detailed clinical phenotyping here allows a more comprehensive clinical comparison

to be made with the previously reported cases, confirming the presence of a variable pattern of dysmorphic features associated with SCAPER syndrome. It is now clear that the cardinal clinical features of the disorder include intellectual disability and developmental delay particularly affecting speech and language and motor milestones. Intellectual disability in particular appears to be a remarkably penetrant feature, and there has only been a single report of a single individual presenting with non-syndromic retinitis pigmentosa and no evidence of intellectual disability. This individual was found to be homozygous for a c.2023-2A>G *SCAPER* variant (343), with the same variant also reported in homozygous form in two further patients with retinitis pigmentosa, ADHD and mild intellectual disability (341). For the SCAPER patient with apparent non-syndromic retinitis pigmentosa, collateral history from parents and the paediatrician indicated that all developmental milestones had been met in time with no signs or symptoms suggestive of intellectual disability, although a formal evaluation did not appear to have been performed. There can sometimes be difficulties in conclusively defining milder developmental delay where more subtle clinical findings may not be identified if not specifically assessed. As such, it is still unclear if intellectual disability is invariably present or is instead a common but variably penetrant feature of the SCAPER syndrome.

Early adult-onset retinitis pigmentosa is another key clinical finding, and the retinal phenotype appears remarkably consistent. In all individuals for whom data are available, progressive loss of night vision begins in first or second decade of life. Together with studies in mice demonstrating expression of SCAPER in multiple retinal layers, particularly in the RPE and photoreceptor inner and outer segments, this supports a role for SCAPER in photoreceptor function and/or maintenance (341).

Hyperactivity appears to be a common feature, with some affected individuals receiving a formal diagnosis of ADHD. Additionally, variable forms of dysmorphic facies are noted in eight individuals in four families, and include a prominent nose, prominent maxilla, narrow chin or micrognathia, high forehead or frontal bossing, and almond-shaped eyes.

Tapering fingers, brachydactyly and proximally placed thumbs, described in eight individuals from two consanguineous Bedouin families belonging to the same tribe in

southern Israel, were also identified as a consistent feature in the two Amish siblings, confirming the association of this feature with the SCAPER syndrome. Short stature and obesity were also a common feature amongst the affected Bedouin patients, and this constellation of clinical features including retinitis pigmentosa, obesity, short stature, intellectual disability, developmental delay and brachydactyly has consequently led to a suggested diagnosis of BBS in these individuals. Preliminary functional studies have demonstrated localisation of SCAPER to primary cilia when over-expressed in cell culture, as well as SCAPER mutations affecting length of cilia in patient fibroblasts, suggestive of a possible role of SCAPER in ciliary dynamics and disassembly (342). Although there is some overlap between the clinical features characteristic of ciliopathies and those seen in SCAPER syndrome, the Amish siblings (who are of normal height and weight for age) demonstrate that the digital, retinal and cognitive abnormalities may occur independently of short stature and obesity. The other common primary features of BBS, including renal anomalies, post-axial polydactyly, hypogonadism (in males) and genital abnormalities (in females) (300), have not been reported in association with *SCAPER* mutation. Whether SCAPER is indeed a ciliary protein contributing to a ciliopathy syndrome still remains unclear.

The inverted nipples and pes planus, noted on examination of both Amish siblings, have not been previously noted in other individuals with *SCAPER* variants. These features may be unique to this family due to inherited autozygosity or may be incompletely expressed in patients with *SCAPER* variants. There appears to be significant variability in the presence and severity of extraocular features associated with the SCAPER syndrome (Table 4.4), and detailed assessment of further patients with respect to these features may be helpful in clarifying their possible association.

Associated ocular pathology may remain undetected or unrecognised in individuals with intellectual disability, as such individuals often have difficulty recognising or articulating their visual symptoms. This highlights the importance of visual screening and ophthalmological assessment in these patients. Other common ocular features include cataracts [in particular posterior subcapsular cataracts, which are commonly associated with retinitis pigmentosa (351)], strabismus, and (high) myopia, with nystagmus and keratoconus noted in a single patient. The high incidence of cataracts, a potentially treatable cause of sight loss, as well as the association with high myopia

and its risk of ocular complications, again supports the case for ophthalmological screening in early childhood.

The allele frequency (~0.04) of the Ohio Amish *SCAPER* c.2236dupA; p.(Ile746Asnfs*6) founder mutation suggests that despite no previous reports, this disorder likely represents an under-recognised cause of retinitis pigmentosa and mild intellectual disability within this community. This highlights the importance of careful clinical evaluation in children and adults with intellectual disability and enables targeted genetic testing for this *SCAPER* variant for Amish individuals with a similar clinical presentation, allowing a more rapid molecular diagnosis and shortening the diagnostic odyssey for affected individuals within the community.

Together with the clinical review of all previously published patients, this study expands the molecular spectrum of disease-causing *SCAPER* variants and enables a clearer delineation of the core (and variable) phenotypical features of SCAPER syndrome to be characterised. These findings also highlight the importance of prompt visual screening and ophthalmic assessment in all individuals with *SCAPER*-associated disease. Despite the increasing numbers of individuals identified with SCAPER syndrome, the precise functions of SCAPER in human growth and development are not fully understood. Further studies to elucidate the precise molecular and developmental roles of this molecule in human growth and skeletal, retinal and brain development and function, will yield important insights into the clinical heterogeneity increasingly observed in *SCAPER*-associated disease.

4.5 Conclusions and future work

Four Amish and Pakistani families with overlapping clinical features suggestive of a ciliopathy disorder underwent genetic investigations in order to provide an accurate molecular diagnosis which could guide the clinical management of affected individuals. These studies clearly demonstrate the utility of genomic testing in clarifying the precise diagnosis for individuals with ciliopathy disorders.

This study reports on only the second MORM syndrome family described in literature, and consolidates the *INPP5E* c.1879C>T; p.(Gln627*) likely Pakistani founder mutation as the cause of this condition. *INPP5E* variants are more commonly associated with the ciliopathy JBTS, and although MORM syndrome appears to be a phenotypically distinct disease entity, it is still unclear if MORM syndrome should instead be considered within the JBTS clinical spectrum. Neuroimaging studies in MORM syndrome patients to confirm the presence or absence of the “molar tooth” sign characteristic of JBTS would be important to clarify this. Functional studies have described the impact of the MORM variant on INPP5E protein interaction, ciliary localisation and catalytic activity (251). The generation of MORM and JBTS mutation-specific murine and other animal models will be useful to further characterise any genotype-phenotype correlation, and provide a useful resource for various cellular and molecular studies to investigate the pathomolecular basis of disease. Given that the MORM syndrome-associated variant is a nonsense variant, TRIDs may be a valid therapeutic approach to explore. These drugs have already been assessed in some ciliopathies including Usher syndrome (352) and retinitis pigmentosa (353).

Characterisation of the *BBS1* p.(Met390Arg) variant in an extended Amish family confirms the phenotypic heterogeneity associated with homozygosity for this recurrent variant, although it did not find the evidence to either support or refute the complex digenic triallelic inheritance hypothesis in BBS. This hypothesis should be evaluated further, in order to provide accurate recurrence-risk estimates to at-risk family members, and could aid in the understanding of bio-molecular pathways involved in BBS-associated phenotypes, as well as facilitating the understanding of complex inheritance in other genetic disorders. Zebrafish and mouse models have been widely used in studies of ciliary dysfunction and ciliopathies including BBS (354, 355), with a

homozygous hypomorphic *Cep290* mutant mouse model shown to demonstrate a more severe phenotype associated with the additional loss of *Bbs4* alleles (317). Similar studies of other BBS loci in the *Bbs1* Met390Arg knock-in mouse (321) and other BBS animal models may help clarify the role of modifier genes in BBS phenotypic expression. Further studies involving the systematic genetic screening of multiple ciliopathy genes combined with detailed longitudinal phenotyping in large ciliopathy patient cohorts will also be useful to define the interaction and true effects of modifier loci, and clarify the molecular basis and disease pathways underlying the clinical complexity in BBS and other ciliopathies. Retinal degeneration is a primary feature of BBS, associated with a poor visual prognosis, with affected individuals often progressing to complete blindness by 15-20 years of age (356). There have been some early promising results from retinal gene replacement therapy studies, showing trends towards improved retinal electrophysiological function following AAV-mediated subretinal gene delivery in *Bbs4*-null and homozygous *Bbs1*$^{M390R/M390R}$ mice (357, 358). This could potentially be combined with an antiapoptotic therapy such as systemic Tauroursodeoxycholic acid (TUDCA), which has previously been shown to slow retinal degeneration in *Bbs1*$^{M390R/M390R}$ mice (359). Given that the majority of BBS-associated genes code for components of the core BBSome complex or chaperonin complex, one major challenge of gene replacement therapy however is the need to avoid overexpression of the gene product which could disrupt protein complex stoichiometry or lead to cell toxicity (357).

The novel *SCAPER* c.2236dupA; p.(Ile746Asnfs*6) founder variant identified within the Ohio Amish population significantly expands the molecular spectrum of disease-causing *SCAPER* variants, and detailed phenotyping studies in our patient cohort enabled a clearer delineation of the core (and variable) phenotypical features of SCAPER syndrome. Although *SCAPER* is now considered by some to be a BBS-associated disease gene (301), it is still unclear whether it is indeed a ciliary protein. Preliminary functional studies have demonstrated *SCAPER* localisation to primary cilia when over-expressed in cell culture. It should be noted however that protein over-expression in cells, particularly with a large tag that may affect protein folding, may lead to altered localisation and may not always reflect the true endogenous location of the protein (342). Additionally, a review of all published SCAPER syndrome patients finds that other common primary features of ciliopathies, including obesity, renal

anomalies, post-axial polydactyly, hypogonadism (in males) and genital abnormalities (in females) are not always present in affected individuals with SCAPER syndrome. Further studies are needed to elucidate possible roles of *SCAPER* in primary cilia function and dynamics. In the adult mouse retina, SCAPER localises to multiple retinal layers, with highest expression observed in the RPE and the photoreceptor outer and inner segments, which are terminally differentiated postmitotic cells (341). The retinal phenotype in SCAPER syndrome does not appear to be congenital in onset, instead presenting in the first or second decade of life. *SCAPER* does not therefore appear to play a developmental role in the retina, and may instead be important for photoreceptor function and/or maintenance. SCAPER function in the retina does not appear to be related to its interaction with cyclin A and its putative role in cell cycle control. Further studies to identify possible SCAPER interacting partners in the retina may facilitate a better understanding of the underlying molecular disease pathways, thereby allowing the development and early deployment of therapies to halt the disease process prior to photoreceptor cell death.

5 IMPROVING KNOWLEDGE OF THE SPECTRUM AND CAUSES OF RARE AND ULTRA-RARE GENETIC EYE DISEASES IN COMMUNITIES

5.1 Introduction

Definitions of rare and ultra-rare disorders vary by country and are not well defined, depending on factors such as disease prevalence and severity, heritability and treatment availability (360). In Europe, rare disorders are defined as diseases or conditions that affect fewer than 1 in 2000 people in the population, whilst ultra-rare disorders generally refer to disorders affecting fewer than 1 in 50,000 people (or fewer than 20 patients per million) (360, 361). It is currently estimated that there are over 6000 rare disorders (362), with new conditions continually being identified as research advances. Although individually these disorders are uncommon, collectively they can affect up to 30 million individuals in Europe (363) and 300 million people worldwide (361).

Over 70% of rare diseases have an identified genetic origin (362); in fact, many genetic disorders have a relatively low prevalence and incidence and are classified as rare or ultra-rare disorders (364). Rare and ultra-rare disorders are often chronic conditions associated with significant morbidity, and are also commonly difficult to diagnose and treat. In fact, these conditions are often categorised as orphan diseases to highlight their severity and the insufficient resources and knowledge available to develop diagnostic assays or treatments.

There has been tremendous interest and advances in the field of rare and ultra-rare genetic eye diseases in recent times, with increasing characterisation of causal genes and underlying disease pathways and delineation of novel inheritance models (365). New therapeutic approaches are entering the clinical phase of development, and the first ever licensed gene replacement therapy for the treatment of a genetic disease in humans, Luxturna® (voretigene neparvovec), was recently approved for the treatment of patients with *RPE65*-associated retinal dystrophy (366).

Genetic eye diseases are however extremely difficult to study in a global setting due to their rarity and associated phenotypic, genetic and allelic heterogeneity. The study of inherited eye diseases in genetically isolated populations such as rural Palestinian or Pakistani communities provides an important opportunity to learn about the genetic causes of inherited eye diseases, whilst also providing desperately required healthcare benefits for the families and populations involved. An improved understanding of the underlying genetic and molecular causes of disease arising from the identification of novel disease genes and variants within communities may then be applied to genetically diverse populations such as the UK, and bring about clinical benefits for patients worldwide.

In this chapter, clinical and genomic investigations were undertaken in families from Pakistani and Palestinian communities with a preliminary clinical diagnosis of various forms of genetic eye diseases. These studies identified 18 likely disease-causing variants in these families, leading to an accurate molecular diagnosis and subsequent refinement of the initial clinical diagnosis, with important treatment and prognostic implications for affected individuals and their families.

Within this chapter, I was responsible for the interpretation and analysis of all collected clinical data for all affected individuals (families 49 - 66). I performed DNA extraction for a proportion of affected families (remaining DNA extraction largely completed by Joe Leslie, University of Exeter). I was also responsible for the analysis of all exome sequencing results, as well as primer design and subsequent cosegregation studies for all variants identified in families within chapters 5.2 and 5.3 (apart from families 50 – 51, which performed in collaboration with Dr Shamim Saleha, Kohat University of Science and Technology, Pakistan). I performed the literature reviews described in this chapter (apart from review of all disease-causing *ALDH1A3* variants, which was again performed in collaboration with Dr Shamim Saleha).

5.2 Consolidating biallelic *SDHD* variants as a cause of mitochondrial complex II deficiency

5.2.1 Introduction

The mitochondrial oxidative phosphorylation (OXPHOS) system is composed of five multi-subunit transmembrane protein complexes (I-V) encoded by the mitochondrial and nuclear genomes, and is the primary mechanism for adenosine triphosphate (ATP) production in eukaryotic cells. OXPHOS defects result in mitochondrial disease, with an estimated prevalence of 1:4300 (367, 368).

Mitochondrial complex II (succinate dehydrogenase) is composed of two catalytic subunits (SDHA/SDHB) anchored to the inner mitochondrial membrane by two smaller subunits (SDHC/SDHD) (369, 370). Complex II differs from other mitochondrial respiratory chain complexes, in that its four structural subunits and their two assembly factors (*SDHAF1*/*SDHAF2*) are solely encoded by the nuclear genome. Complex II is also unique in being involved in both the mitochondrial respiratory chain and the Krebs cycle (368).

Mitochondrial complex II deficiency with multisystem involvement has been reported in association with biallelic *SDHA* (371), *SDHB* (368)*, SDHD* (372, 373) and *SDHAF1* (369, 374) gene variants, with clinical presentations including Leigh syndrome, leukoencephalopathy, optic atrophy and cardiomyopathy with highly variable severity and age of onset (371, 375). Complex II deficiency is rare, accounting for only 2-4% of OXPHOS defects (372), with variants in *SDHA* being most common, predominantly associated with Leigh syndrome (371). Previously, only two individuals with candidate biallelic *SDHD* variants and isolated complex II deficiency have been reported (372, 373). This study describes four Palestinian siblings presenting in childhood with clinical features indicative of mitochondrial disease and a likely pathogenic homozygous *SDHD* variant, consolidating *SDHD* gene variants as a likely cause of autosomal recessive mitochondrial complex II deficiency.

5.2.2 Materials and methods

Blood samples were collected with informed consent for DNA extraction (see section 2.3.2). SNP genotyping was performed in affected individuals V:2, V:4 and V:6 as described in section 2.3.4. WES was undertaken using DNA from a single affected individual (V:4) at BGI Hong Kong, as described in section 2.3.5. Primer design, PCR and dideoxy sequencing (Appendix Table D2) was performed as described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals.

5.2.3 Results: clinical and genetic findings

Four affected Palestinian patients (three male, one female) aged 4 - 20 years, comprising of two sibships from an extended interconnecting family (family 49) (Figure 5.1A), were investigated in this study. All four children presented with developmental delay in infancy and variable clinical and laboratory findings suggestive of a mitochondrial disorder including elevated serum lactate/urinary Krebs cycle metabolites, nystagmus, optic atrophy, progressive microcephaly, generalised hypotonia, epileptic seizures, severe/profound intellectual disability/developmental impairment, and cardiomyopathy. The affected children were not dysmorphic (Figure 5.1C), though individuals V:2 and V:4 were noted to have significant hypertrichosis, particularly over their back and limbs. MRI neuroimaging was unremarkable for one child at 8 months (V:2; Figure 5.1D), however, his sister's scan revealed delayed myelination at age 6 months (V:4). Hirschsprung disease, confirmed by aganglionic rectal biopsy, was noted in a single individual (V:2). A full description of the clinical features and disease progression are summarised in Table 5.1.

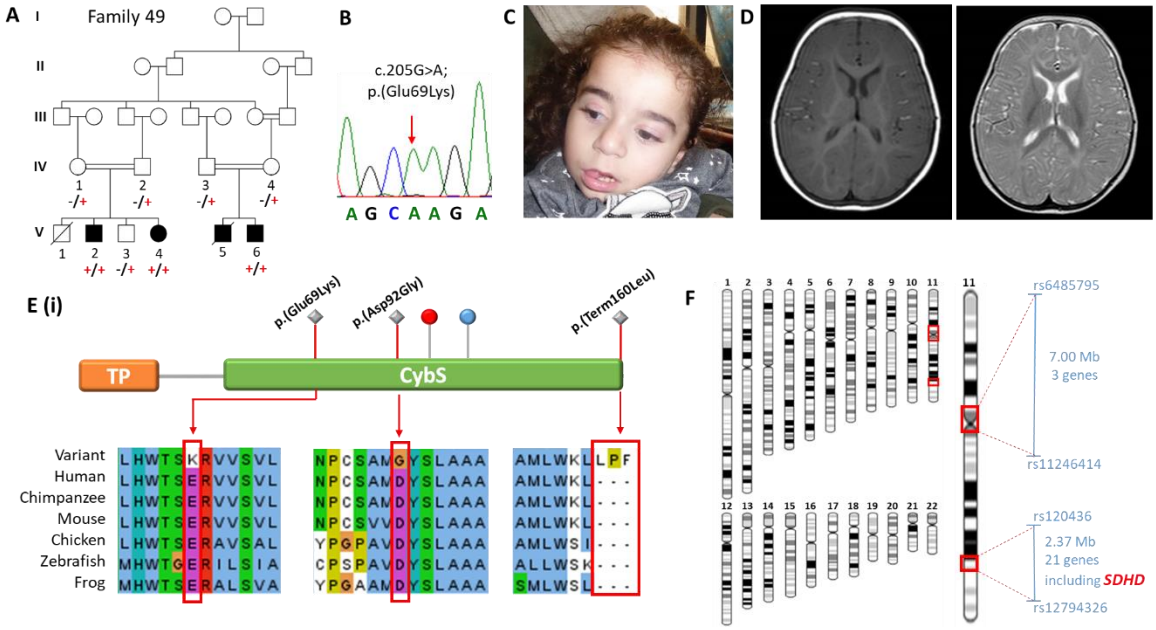


Figure 5.1 Family 49 pedigree showing *SDHD* c.205G>A genotype data, neuroimaging and images of affected individuals

(A) Pedigree diagram of family 49 showing segregation of the SDHD *c.205G>A; p.(Glu69Lys) variant, co-segregating appropriately for an autosomal recessive condition. Genotypes are shown beneath generations IV and V (+, c.205G>A; -, wild type). DNA was available from all but one affected individual (V:5).*
(B) Electropherogram showing the DNA sequence at the position of SDHD *c.205G>A in a homozygous affected individual.*
(C) T1 and T2-weighted axial views of MRI head of individual V:2 (aged 8 months), showing normal myelination and no intracranial abnormalities.
(D) Image of affected individual V:4, illustrating the absence of any facial dysmorphism.
(E)(i) Schematic showing domain architecture of SDHD [adapted from UniProt (216)] and the location of p.(Glu69Lys), p.(Asp92Gly) and p.(Ter160LeuextTer3) variants within the succinate dehydrogenase cytochrome b small subunit (CybS) domain. The orange rectangle denotes the transit peptide (TP) domain, the red circle denotes the iron (heme axial ligand) binding site shared with SDHC, and the blue circle denotes the ubiquinone binding site shared with SDHB (ii) Conservation analyses: multiple species alignments of amino acid sequences of SDHD at the locations of the p.(Glu69Lys), p.(Asp92Gly) and p.(Ter160LeuextTer3) variants.
(F) Visual depiction of the two autozygous regions on chromosome 11 (shown in red) common to affected individuals V:2, V:4 and V:6 including the 2.37Mb region containing 21 genes including SDHD.

Table 5.1 Clinical features of affected individuals with mitochondrial complex II deficiency due to biallelic *SDHD* variants

	Jackson *et al*	Alston *et al*	This study; V:2	This study; V:4	This study; V:6	This study; V:5
Genotype (NM_003002.3)	p.(Glu69Lys); (Ter160LeuextTer3)	p.(Asp92Gly); (Asp92Gly)	p.(Glu69Lys); (Glu69Lys)	p.(Glu69Lys); (Glu69Lys)	p.(Glu69Lys); (Glu69Lys)	NA (deceased)
Ethnicity	Swiss	Irish	Palestinian	Palestinian	Palestinian	Palestinian
Gender	F	M	M	F	M	M
Age at last evaluation	7 yrs (Deceased age 10 yrs)	Deceased day 1 of life from lethal cardiomyopathy	6.4 yrs	4.5 yrs	20 yrs	**Hx from parents** Deceased age 10 yrs - cardiac arrest
GROWTH PARAMETERS						
Birth weight kg (SDS)	NA	2.62	3.5 (-0.1)	2.8 (-1.4)	NA	NA
Birth OFC cm (SDS)	NA	34.5	35 (-0.2)	35 (+0.4)	NA	NA
OFC cm (SDS)	2° microcephaly from 2 yrs	NA	46.5 (-4.4)	49 (-2.0)	NA	NA
DEVELOPMENT						
Intellectual disability	Severe Maximum developmental age of 11 mo at 4 yrs	NA	Profound	Profound	Severe	Severe
Developmental regression	✓ (from age 3 mo after bronchiolitis) Several subsequent episodes of regression after infection/prolonged fasting	NA	✓ (from age 5 mo following surgery) Previously was sitting with support, mouthing, purposeful hand movements	✓ (from age 4 mo) Previously sitting with support, fixing and following, mouthing, purposeful hand movements	No clear hx of regression	No clear hx of regression
Gross motor skills	NA	NA	Antigravity movements of arms and legs only	Sits with support. Rolls from back to front	Sits with support	Walked with support
Fine motor skills	NA	NA	No active hand use	No active hand use	Finger feed	Finger fed
Expressive and receptive language development	NA	NA	Vocalisation, Makes sucking motions if thirsty	Vocalisation, Responds to loud noises	Vocalisation Points to indicate needs	2 word phrases
Behavioural abnormalities	NA	NA	Sleep disturbances treated with Risperidone	None	Repetitive hand movements	-
NEUROLOGY						
Generalised muscle hypotonia	✓	NA	✓	✓	✓	NK

Movement disorder	Dystonia and ataxia	NA	NA	NA	Dystonia	-
Seizures	✓ Polymorphic epilepsy and intractable myoclonic movements	NA	Generalised seizures post surgery (6 mo)	Abnormal movements (4 mo)	Seizures when younger now resolved	-
EEG	Normal	NA	Normal (7 mo)	Normal	NK	NK
Neuroimaging	Normal MRI (10 mo and 2 yrs)	NA	Normal MRI (8 mo)	MRI: Delayed myelination (6 mo)	Normal CT brain (7 yrs)	NK
OCULAR						
Visual impairment	✓	NK	✓	✓	✓	NK
Nystagmus	✓	NK	✓	✓	✓	✓
Optic atrophy	✓	NK	✓	✓	✓	NK
Strabismus	NA	NK	NK	✗	✓	✓
HEARING IMPAIRMENT	✗	NK	✗	✗	✗	✗
CARDIAC ABNORMALITIES	NK	✓ Hypertrophic cardiomyopathy with Lt ventricular non-compaction (prenatal onset)	✗ normal cardiac structure and function (7 yrs)	✗ normal cardiac structure and function (2.8 yrs)	✓ Minimal Lt ventricular hypertrophy, with low normal Lt ventricular function (21yrs)	✓ Echo (5 yrs): Dilated cardiomyopathy with reduced left ventricular ejection fraction, mild mitral and tricuspid regurgitation
HYPERTRICHOSIS	NK	NK	✓	✓	✗	NK
METABOLIC INVESTIGATIONS	Raised serum lactate (10.2 mmol/L), Lactic aciduria and ketonuria, urinary Krebs cycle intermediates Marked defect in complex II activity in muscle homogenate	Marked defect in complex II activity in muscle homogenate	None	Raised serum lactate (5.58 mmol/L) Urinary excretion of Krebs cycle metabolites (succinic, fumeric and ketoglutauric acids)	Normal respiratory chain complexes II-IV in fibroblast homogenate (succinate: cytochrome c reductase assay was outside the normal range, but reported as normal) Non-specific muscle biopsy findings	NK
OTHER CLINICAL FEATURES	-	-	Hirschsprung disease diagnosed at 1 mo, Frequent LRTI	-	-	-

Abbreviations: CT, computerised tomography; Echo, echocardiography, F, female; hx, history; M, male; mo, months; MRI, magnetic resonance imaging; NA, not available; NK, not known; OFC, occipitofrontal circumference; SDS, standard deviation scores; LRTI, lower respiratory tract infection; Lt, left; yr, years; the (✓) and (✗) symbols indicate the presence or absence of a feature in an affected subject respectively

Height, weight, BMI and OFC Z-scores were calculated using a Microsoft Excel add-in to access growth references based on the LMS method (349) using a reference European population (350)

Genome-wide SNP genotyping and WES was undertaken assuming that a homozygous founder variant was responsible, although also considering other genetic mechanisms. SNP genotyping (individuals V:2, V:4 and V:6) identified four notable (>1 Mb) shared homozygous regions, the two largest identified on chromosome 11; a ~7.00 Mb region (rs6485795 – rs11246414, chr11:g.47908294 – 54905443 [hg38]), and a ~2.42 Mb region (rs120436 – rs12794326, chr11:g.110826521 – 113248134) (Figure 5.1F).

WES was performed in affected individual V:4 in family 49 to identify rare functional candidate variants. Variants were prioritised as described in section 2.3.5 and cross referenced with SNP data, identifying only a single candidate homozygous variant of relevance to the phenotype in *SDHD* [NM_003002.3:c.205G>A; p.(Glu69Lys); chr11:g.112088902G>A], located within the second largest shared homozygous region. This variant is present in only two heterozygotes in gnomAD (v2.1.1) and is predicted to result in a glutamic acid to lysine substitution in an evolutionarily conserved Glu69 residue (Figure 5.1E). This variant was previously reported as the likely candidate cause of disease in compound heterozygous form in a single individual with autosomal recessive encephalomyopathy and isolated mitochondrial complex II deficiency (ClinVar accession: VCV000156153.6 and SCV001424558) (372). Dideoxy sequencing confirmed co-segregation as appropriate for an autosomal recessive disorder (Figure 5.1A/B).

5.2.4 Discussion

A homozygous *SDHD* c.205G>A; p.(Glu69Lys) missense variant was identified as the likely cause of isolated mitochondrial complex II deficiency in three affected children from an extended Palestinian family in this study. DNA was unavailable for individual V:5, (deceased age 10 years), whose clinical history overlapped that of his sibling (individual V:6). Tissues and organs heavily dependent on robust oxidative phosphorylation processes tend to be most affected by mitochondrial disease (376), explaining why common findings include optic atrophy, leukoencephalopathy, myopathy, cardiomyopathy and Leigh syndrome. These clinical features overlap those described in the affected individuals in this study, as well as the two individuals with *SDHD*-related mitochondrial disease reported to date (Table 5.1).

Previously, compound heterozygous variants in *SDHD* (372) including the same p.(Glu69Lys) variant identified here and a c.479G>T; p.(Ter160LeuextTer3) alteration were identified as the likely candidate cause of disease in a Swiss child presenting with developmental regression following a viral infection at three months. Progressive ocular (visual impairment, nystagmus, optic disc pallor) and neurological (epileptic seizures, ataxia, dystonia and continuous intractable myoclonic movement) involvement were described, and the child died aged 10 years. Urinalysis revealed lactic aciduria, ketonuria and Krebs cycle intermediates. Complex II activity was deficient in skeletal muscle, and complementation studies in patient fibroblasts showed restoration of complex II assembly and function with expression of wild-type, but not mutant, *SDHD* cDNA (372). Subsequently an Irish male infant was described (373) homozygous for a novel *SDHD* c.275A>G; p.(Asp92Gly) substitution, presenting with cardiomyopathy *in utero*. He developed cardiopulmonary insufficiency rapidly after birth, dying on day 1 of life. Subsequent analysis of respiratory chain function in patient muscle homogenate revealed a marked defect in complex II activity. The four affected individuals described here show phenotypic overlap with both these individuals (Table 5.1).

This study extends the clinical spectrum and highlights the wide range of phenotypical features and severity across affected individuals, even those with the same *SDHD* genotype (Table 5.1). Hypertrichosis, a recognised feature of some forms of mitochondrial disease (most notably *SURF1*-associated Leigh syndrome) (377), was a noted feature in two Palestinian children. Hirschsprung disease diagnosed in a single affected individual (V:2) has not been previously reported in association with *SDHD* variants, and it remains unclear whether this is an associated or unrelated feature. Neurodevelopmental regression is a common characteristic of mitochondrial disease, particularly during physiologic stress through intercurrent infection, prolonged fasting or dehydration (378). It is thus unsurprising that this appears to be a common feature of complex II deficiency due to biallelic *SDHD* variants (Table 5.1). An accurate molecular diagnosis for complex II deficient patients would support avoidance of prolonged fasting and dehydration.

In addition to their role in primary mitochondrial disease, heterozygous germline variants in *SDHD* and other complex II subunits and assembly factors (including *SDHA*, *SDHB*, *SDHC* and *SDHAF2)* are associated with paragangliomas, phaeochromocytomas and gastrointestinal stromal tumours (369). Dominantly inherited hereditary cancer-associated *SDHD* variants exhibit a parent of origin effect; typically only paternally inherited mutations are associated with disease (379, 380). This was originally interpreted as evidence for sex-specific epigenetic modification of the maternal *SDHD* allele (381); however, the *SDHD* gene does not appear to be imprinted, being biallelically expressed in human brain, kidney and lymphoid tissue (379). Instead, the 'Hensen Model' suggests that tumour development only occurs following loss of function of both copies of *SDHD* as well as loss of an imprinted (paternally silenced and maternally active) tumour suppressor gene likely found within the chromosome 11p15 region, which contains several loci regulated by genomic imprinting as well as encompassing the candidate tumour suppressor genes *HK19* and *CDKN1C* (382, 383).

None of the three *SDHD* variants associated with mitochondrial complex II deficiency have been previously linked to tumourigenesis, including in this extended Palestinian family, although a Dutch founder familial paraganglioma *SDHD* variant c.274G>T; p.(Asp92Tyr) affecting the same Asp92 amino acid residue as the p.(Asp92Gly) mitochondrial complex II deficiency-associated variant has been described (384). Additionally, *SDHA* and *SDHB* variants have been associated with both mitochondrial complex II deficiency in biallelic form, and hereditary cancer susceptibility in monoallelic form (371, 385). Therefore, routine surveillance of heterozygous *SDHD* carriers is suggested for early detection of paragangliomas and phaeochromocytomas facilitating appropriate intervention.

Together the data presented here consolidate biallelic *SDHD* variants as a cause of mitochondrial disease due to mitochondrial complex II malfunction, and extend the variable clinical features associated with the condition.

5.3 Informing clinical care through genomic studies in Pakistani families with inherited ocular diseases

5.3.1 Introduction

In Pakistan, blindness presents a significant public health issue with an estimated prevalence of 15 per 10,000 in children under the age of 15 years (386), and inherited eye diseases are an important contributor to this burden (183, 387). Due to regional geographical limitations or restricted local clinical resources, however, there is often limited access to ophthalmic equipment and expertise needed for performing and interpreting the detailed phenotyping studies needed for an accurate clinical diagnosis. In a setting where clinical information is limited, the application of next generation sequencing approaches may enable an accurate molecular diagnosis to be achieved, providing important clinical benefits to affected individuals and their families, as well as providing scientific insights into the genomic architecture of inherited eye diseases within local communities as well as globally.

As part of an ongoing international collaboration, genomic studies were undertaken in 17 families from Pakistan with a preliminary clinical diagnosis of anophthalmia, congenital cataracts, congenital blindness, nystagmus and OCA. This study identified 14 likely disease-causing mutations in these families, leading to an accurate molecular diagnosis and subsequent refinement of the initial clinical diagnosis, with important treatment and prognostic implications for affected individuals and their families.

5.3.2 Materials and methods

Genetic and clinical investigations in 17 Pakistani families were undertaken with informed consent. Medical histories were taken to enable the initial clinical diagnosis. Following results of genetic analyses and identification of likely pathogenic variants, affected individuals were revisited, and a detailed medical history including documentation of symptoms was ascertained. Clinical features were documented with facial photographs and videos, as well as external ocular photographs. Visual acuity testing using Snellen charts, anterior and posterior segment examinations on a slit-

lamp biomicroscope and B-scan ultrasonography was performed in a limited number of affected individuals.

Blood sample collection and DNA extraction was performed as previously described (see section 2.3.2

2.3.2 DNA extraction and quantification). WES (Exeter Sequencing Service, University of Exeter, UK) (families 50 - 52 and 59), or Illumina TruSight™ One clinical exome sequencing panel (families 56 - 58, 60 - 66) was performed as described in section 2.3.5 in a single affected individual in each family (apart from family 59, where two affected individuals underwent WES). Bioinformatic analysis of exome data was performed as per section 2.3.5. Primer design, PCR and dideoxy sequencing (Appendix Tables D1 and D2) was performed as described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals.

A literature review was performed as described in section 2.4 to retrieve all reported *ALDH1A3*, *TDRD7* and *ATOH7* inherited ocular disease-associated variants. Findings are summarised in Tables 5.6 – 5.8.

5.3.3 Results: clinical and genetic findings

Study subjects from 17 families with individuals diagnosed with ocular conditions alongside parents and unaffected siblings were enrolled from different provinces in Pakistan (Sindh, KPK, Punjab and Balochistan). Two families were initially diagnosed with anophthalmia (families 50 and 51) six families with non-syndromic congenital cataracts (families 52 - 55, 57, 58), five families with OCA (families 62 - 66), one family with congenital blindness (family 59) and three families with nystagmus (families 56, 60, 61). Preliminary clinical findings are summarised in Table 5.2.

Following molecular genetic findings, affected individuals underwent a further ophthalmic examination where possible, allowing confirmation of the initial clinical diagnosis in eight families (families 50 - 56 and family 60). In some individuals, additional ocular pathologies were identified, allowing a refinement of the initial clinical

diagnoses in the remaining nine families (families 57 - 59 and 61 - 66). These results are summarised in Table 5.2.

Table 5.2 Variants responsible for inherited ocular diseases identified in families 50 - 66

Family	Region, province (Caste)	Initial clinical diagnosis	Preliminary ocular findings	Genotype (affected individuals)	Previously reported	gnomAD MAF all/SAS (hom count)	Revised diagnosis	Additional ocular findings post molecular diagnosis
50	KPK	Anophthalmia	Bilateral isolated anophthalmia	***ALDH1A3*** c.1240G>C; p.(Gly414Arg)/ c.1240G>C; p.(Gly414Arg)	Novel	Not present	Isolated anophthalmia	Nil
51	KPK	Anophthalmia	Bilateral isolated anophthalmia	***ALDH1A3*** c.172dupG; p.(Glu58Glyfs*5)/ c.172dupG; p.(Glu58Glyfs*5)	Novel	Not present	Isolated anophthalmia	Nil
52	Agra road, Sindh (Jamro)	Congenital cataract	Bilateral congenital cataracts	***FYCO1*** c.2206C>T; p.(Gln736*)/ c.2206C>T; p.(Gln736*)	Chen[a] *et al* (388); Chen[b] *et al* (389)	0.00001991/ 0.0001633 (not present)	Congenital cataract	Nil
53	Agra road, Sindh (Jamro)	Congenital cataract	Bilateral congenital cataracts	***FYCO1*** c.2206C>T; p.(Gln736*)/ c.2206C>T; p.(Gln736*)	Chen[a] *et al* (388); Chen[b] *et al* (389)	0.00001991/ 0.0001633 (not present)	Congenital cataract	Nil
54	Agra road, Sindh (Jamro)	Congenital cataract	Bilateral congenital cataracts	***FYCO1*** c.2206C>T; p.(Gln736*)/ c.2206C>T; p.(Gln736*)	Chen[a] *et al* (388); Chen[b] *et al* (389)	0.00001991/ 0.0001633 (not present)	Congenital cataract	Nil
55	Agra road, Sindh (Jamro)	Congenital cataract	Bilateral congenital cataracts	***FYCO1*** c.2206C>T; p.(Gln736*)/ c.2206C>T; p.(Gln736*)	Chen[a] *et al* (388); Chen[b] *et al* (389)	0.00001991/ 0.0001633 (not present)	Congenital cataract	Nil
56	KPK (Pakhtun)	Nystagmus	Nystagmus, visual deterioration, previous intraocular surgeries	***TDRD7*** c.2469delG; p.(Asn824Thrfs*27)/ c.2469delG; p.(Asn824Thrfs*27)	Novel	Not present	Congenital cataract	Nil
57	Agra road, Sindh (Jamro)	Congenital cataract	Visual impairment	***CYP1B1*** c.1169G>A; p.(Arg390His)/ c.1169G>A; p.(Arg390His)	Stoilov *et al* (390); Rauf *et al* (391); Sheikh *et al* (392)	0.0001032/ 0.0002942 (not present)	Primary congenital glaucoma	Severe optic neuropathy
58	Gambat, Sindh (Bareejo)	Congenital cataract	Bilateral congenital cataracts	***ATOH7*** c.94delG; p.(Ala32Profs*55)/ c.94delG; p.(Ala32Profs*55)	Novel	Not present	Global ocular developmental defects	Congenital cataract,

								congenital glaucoma, RD
59	Attock city, Punjab (Gujjar)	Congenital blindness	Visual impairment from birth	***LRP5*** c.1076C>G; p.(Thr359Arg)/ c.1076C>G; p.(Thr359Arg)	Novel	Not present	Global ocular develop-mental defects	Nystagmus, microphthalmia, corneal opacity, iris adhesions, cataract
60	KPK (Baloch)	Nystagmus	Congenital nystagmus	***FRMD7*** c.443T>A; p.(Leu148*)/.	Novel	Not present	Congenital idiopathic nystagmus	Nil
61	Quetta, Balochistan (Pirkani, Bravi)	Nystagmus	Congenital nystagmus	***PAX6*** c.718C>T; p.(Arg240*)/ WT	Glaser *et al* (6); Kondo-Saitoj *et al* (393); Abouzeid *et al* (394); Chograni *et al* (395); Primignani *et al* (396); Lin *et al* (397); Syrimis *et al* (398); Cross *et al* (399); Tian *et al* (400)	Not present	Isolated aniridia	Aniridia
62	Balochistan (Baloch)	OCA	Nystagmus, albinism	***HPS1*** c.1397+1G>A/ c.1397+1G>A	Novel	Not present	HPS	Nil
63	Shadra, Punjab (Bhatti Rajpoot)	OCA	Nystagmus, albinism	***HPS1*** c.437C>T; p.(Trp146*)/ c.437C>T; p.(Trp146*)	Novel	Not present	HPS	Nil
64	Rahim Yar Khan, Punjab (Bukhari Sayyed)	OCA	Nystagmus, albinism	***HPS1*** c.972delC; p.(Met325Trpfs*6)/ c.972delC; p.(Met325Trpfs*6)	Oh *et al* (401)	0.0001595/ 0 (not present)	HPS	Nil
65	Swabi, KPK (Pakhtun, Yousafzai)	OCA	Nystagmus, albinism	***HPS1*** c.517C>T; p.(Arg173*)/ c.517C>T; p.(Arg173*)	Wei *et al* (402); Theunissen *et al* (403)	0.000004033/ 0 (not present)	HPS	Nil
66	Peshawar, Pakhtun	OCA	Nystagmus, albinism	***HPS1*** c.2009T>C; p.(Leu670Pro)/ c.2009T>C; p.(Leu670Pro)	Novel	Not present	HPS	Nil

Abbreviations: gnomAD, genome aggregation database (v2.1.1); hom, homozygous; HPS, Hermansky-pudlak syndrome; KPK, Khyber Paktunkhwa; MAF, minor allele frequency; OCA, oculocutaneous albinism; SAS, South Asia; RD, retinal detachment.

Gene transcripts: ALDH1A3 *(NM_000693.3);* FYCO1 *(NM_024513.3);* TDRD7 *(NM_014290.2),* CYP1B1 *(NM_000104.3),* ATOH7 *(NM_145178.3),* LRP5 *(NM_002335.3),* FRMD7 *(NM_194277.2),* PAX6 *(NM_000280.5),* HPS1 *(NM_000195.3).*

Novel variants in ALDH1A3 *associated with autosomal recessive anophthalmia/microphthalmia*

Bilateral isolated anophthalmia was the major clinical finding in all affected individuals in families 50 and 51; affected individuals were all reported to be of normal intelligence with no signs of intellectual disability, and no other neurological and behavioural features were observed. WES performed in a single affected individual in each family (family 50, individual IV:7; family 61, individual II:1) identified homozygosity for novel *ALDH1A3* variants in each family, summarised in Table 5.2.

The novel c.1240G>C ; p.(Gly414Arg) variant identified in family 50 was absent in gnomAD (v2.1.1 and v3.1.1); there was however another genomic variant (c.1240G>A) resulting in the same p.(Gly414Arg) amino acid substitution that is present in only a single heterozygous non-Finnish European individual in gnomAD, with no homozygous individuals for this variant reported. *In silico* analysis of the *ALDH1A3* p.(Gly414Arg) variant using PolyPhen-2, PROVEAN and SIFT all predict pathogenicity. In family 51, the novel c.172dupG; p.(Glu58Glyfs*5) variant is predicted to result in a frameshift followed by a premature stop codon [p.(Glu58Glyfs*5)], leading to loss of function via mRNA-targeted degradation and nonsense-mediated decay. Both *ALDH1A3* variants segregate as expected for an autosomal recessive condition in each family investigated (Figure 5.2).

A FYCO1 *c.2206C>T; p.(Gln736*) likely regional founder gene variant commonly underlies congenital cataracts in Pakistani families*

In family 52, WES in single affected individual (V:7) identified a single homozygous *FYCO1* premature nonsense variant (GRCh38) c.2206C>T; p.(Gln736*) of relevance to the ocular phenotype. This variant has previously been described in association with congenital cataracts in Pakistani families (388, 389). The same *FYCO1* variant was subsequently identified via targeted dideoxy sequencing in a further three congenital cataract families (families 53 - 55) residing in the same region (Agra Road, Sindh) as family 52 (Table 5.2).

This variant segregated as expected for an autosomal recessive condition in all families investigated (Figure 5.3). Further clinical details are summarised in Table 5.3.

Distinct HPS1 *gene variants are associated with albinism in five Pakistani families*

Affected individuals in families 62 - 66 all had a clinical diagnosis of OCA. TruSight™ One clinical exome sequencing of a single affected individual in each family (family 62, individual II:2; family 63, individual IV:1; family 64, individual IV:3; family 65, individual IV:2 and family 66, individual II:2) identified three novel and two previous reported *HPS1* variants, summarised in Table 5.2.

For the novel *HPS1* c.1397+1G>A splice site variant, *in silico* analysis using the bioinformatics tool Human Splicing Finder predicts an effect on splicing via alteration of the wild-type donor site. The novel *HPS1* p.(Leu670Pro) variant alters an evolutionary conserved leucine residue (Figure 5.5), with bioinformatics tools SIFT, PolyPhen-2 and PROVEAN all predicting pathogenicity. Missense variants located within the same Longin 3 domain as the p.(Leu670Pro) variant have previously been reported in association with Hermansky-Pudlak syndrome (HPS) (141, 348, 404-408). The novel *HPS1* p.(Trp146*) nonsense variant is predicted to result in loss of function via mRNA-targeted degradation and nonsense-mediated decay.

All *HPS1* variants segregated as expected for an autosomal recessive condition in all families investigated (Figure 5.5) and were considered likely causative for HPS in all affected individuals.

Defining sequence variants in six families with inherited ocular disease enables refinement of clinical diagnosis

Illumina TruSight™ One clinical exome sequencing (family 56, individual II:2; family 57, individual V:1; family 58, individual IV:1; family 60, individual IV:2; family 61, individual IV:3) and WES (family 59, individuals IV:2 and IV:5) was

performed in six families with preliminary clinical diagnoses of nystagmus, congenital cataracts and congenital blindness. In all families investigated, novel and known likely pathogenic variants in inherited ocular disease genes that were of plausible relevance to the ocular phenotype were identified, allowing further refinement of the initial clinical diagnoses (summarised in Table 5.2).

The novel *TDRD7* c.2469delG; p.(Asn824Thrfs*27) and *ATOH7* c.94delG; p.(Ala32Profs*55) variants are predicted to result in a frameshift. These variants, as well as the novel *FRMD7* c.443T>A, p.(Leu148*) nonsense variant, are all predicted to result in loss of function via mRNA-targeted degradation and nonsense-mediated decay. The novel *LRP5* c.1076C>G; p.(Thr359Arg) missense variant alters an evolutionary conserved threonine residue, with bioinformatics tools SIFT, PolyPhen-2 and PROVEAN all predicting pathogenicity.

Following genetic investigations, a further limited ophthalmic examination was performed in affected individuals from families 58 and 59, with additional clinical findings remaining consistent with the molecular diagnoses obtained (Tables 5.4 and 5.5).

In family 60, the *FRMD7* p.(Leu148*) variant segregated appropriately for an X-linked condition (penetrant in the carrier female III:2). In family 61, the *PAX6* variant was found to occur *de novo*, and was not identified in either unaffected parent. In the remaining families (families 57 - 59), the variants segregated appropriately for an autosomal recessive condition. All identified variants were thought likely to be responsible for the inherited ocular disorder in the affected families investigated (Figure 5.4).

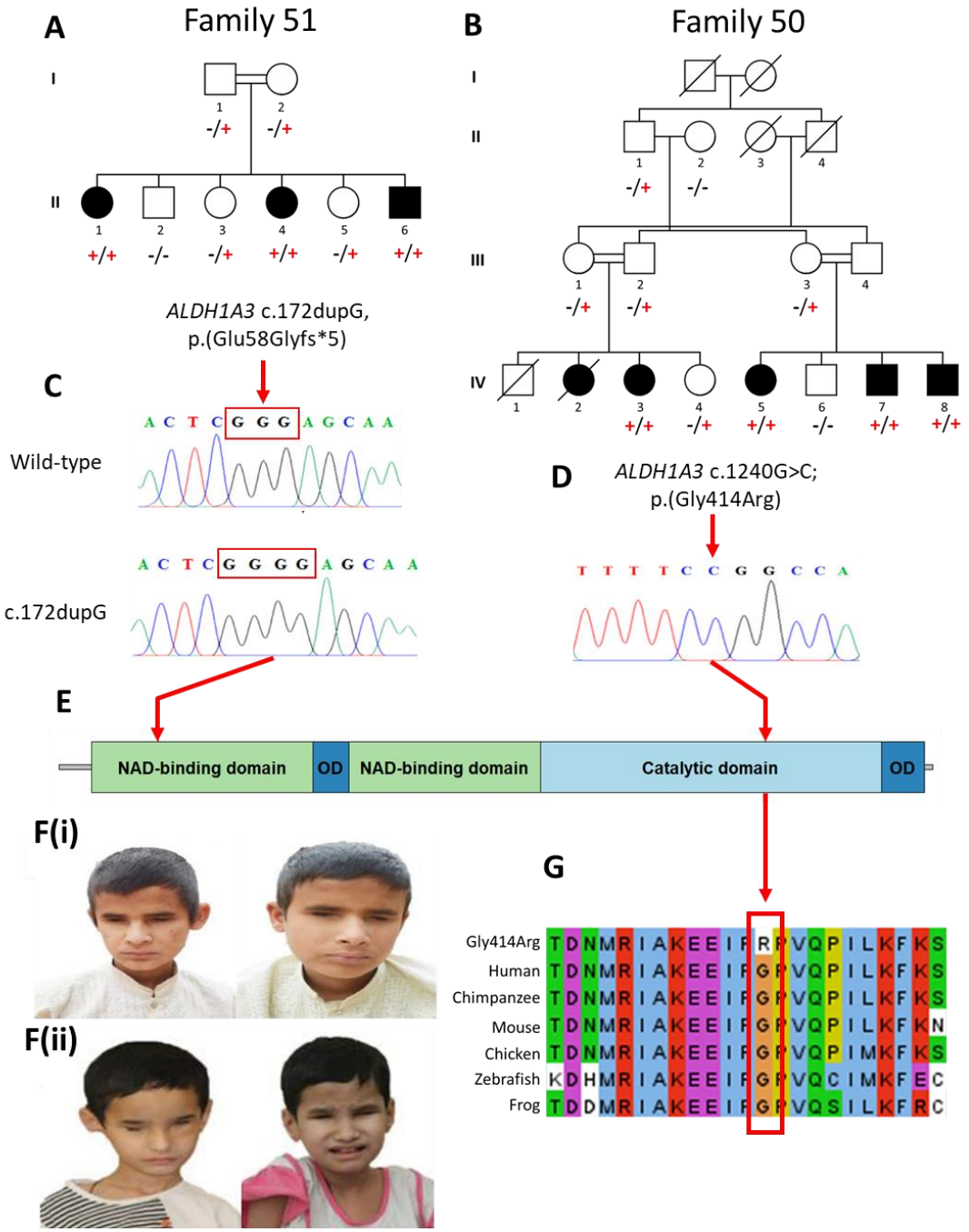


Figure 5.2 Pedigrees and *ALDH1A3* genotype data for families 50 - 51

(A-B) Pedigree diagrams of families 50 and 51 showing segregation of the ALDH1A3 *c.172dupG; p.(Glu58Glyfs*5) and c.1240G>C; p.(Gly414Arg) variants in family 51 (A) and family 50 (B) respectively, co-segregating appropriately for an autosomal recessive condition. Genotypes are shown beneath each family member investigated (+, variant; -, wild type).*
(C-D) Sequence chromatograms of a wild-type individual along with ALDH1A3 *c.172dupG in a homozygous affected individual (C), as well as the DNA sequence at position ALDH1A3 c.1240G>C in a homozygous affected individual (D).*

(E) Schematic showing domain architecture of ALDH1A3 polypeptide [predicted domains described by Moretti et al *(409)] and the location of the c.172dupG; p.(Glu58Glyfs*5) and c.1240G>C; p.(Gly414Arg) variants. OD, oligomerisation domain.*
(F) Photographs of two affected individuals in family 49 [F(i)] and family 50 [F(i)] with non-syndromic clinical anophthalmia.
(G) Conservation analysis: multiple species alignments of partial amino acid sequences of ALDH1A3 showing conservation of Glycine at position 414.

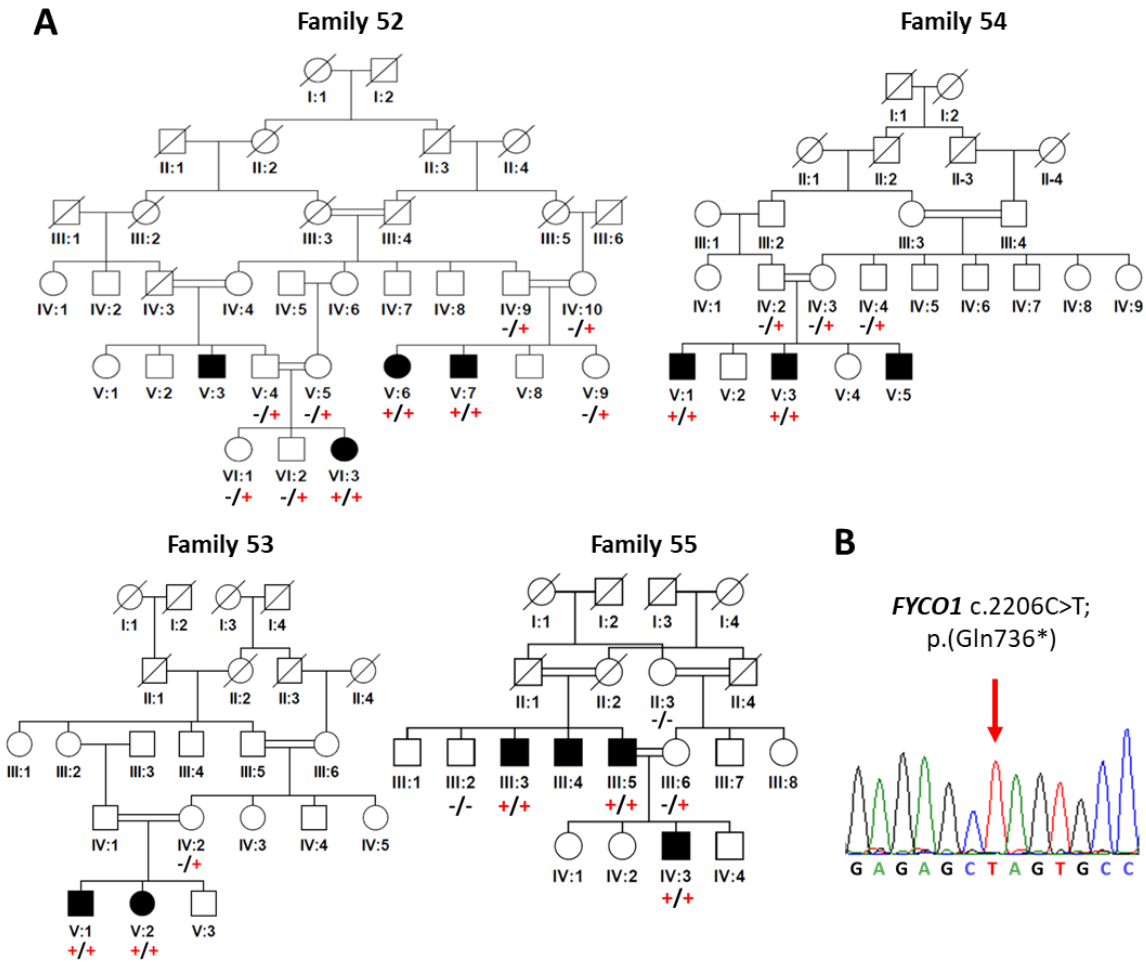


Figure 5.3 Pedigrees and *FYCO1* genotype data for families 52 - 55

(A) Pedigree diagrams in families 52 - 55 *showing segregation of* FYCO1 *c.2206C>T; p.(Gln736*) variants, co-segregating appropriately for an autosomal recessive condition. Genotypes are shown beneath each family member investigated (+, variant; -, wild type).*
(B) Sequence chromatograms of the FYCO1 *c.2206C>T; p.(Gln736*) variant in a homozygous affected individual.*

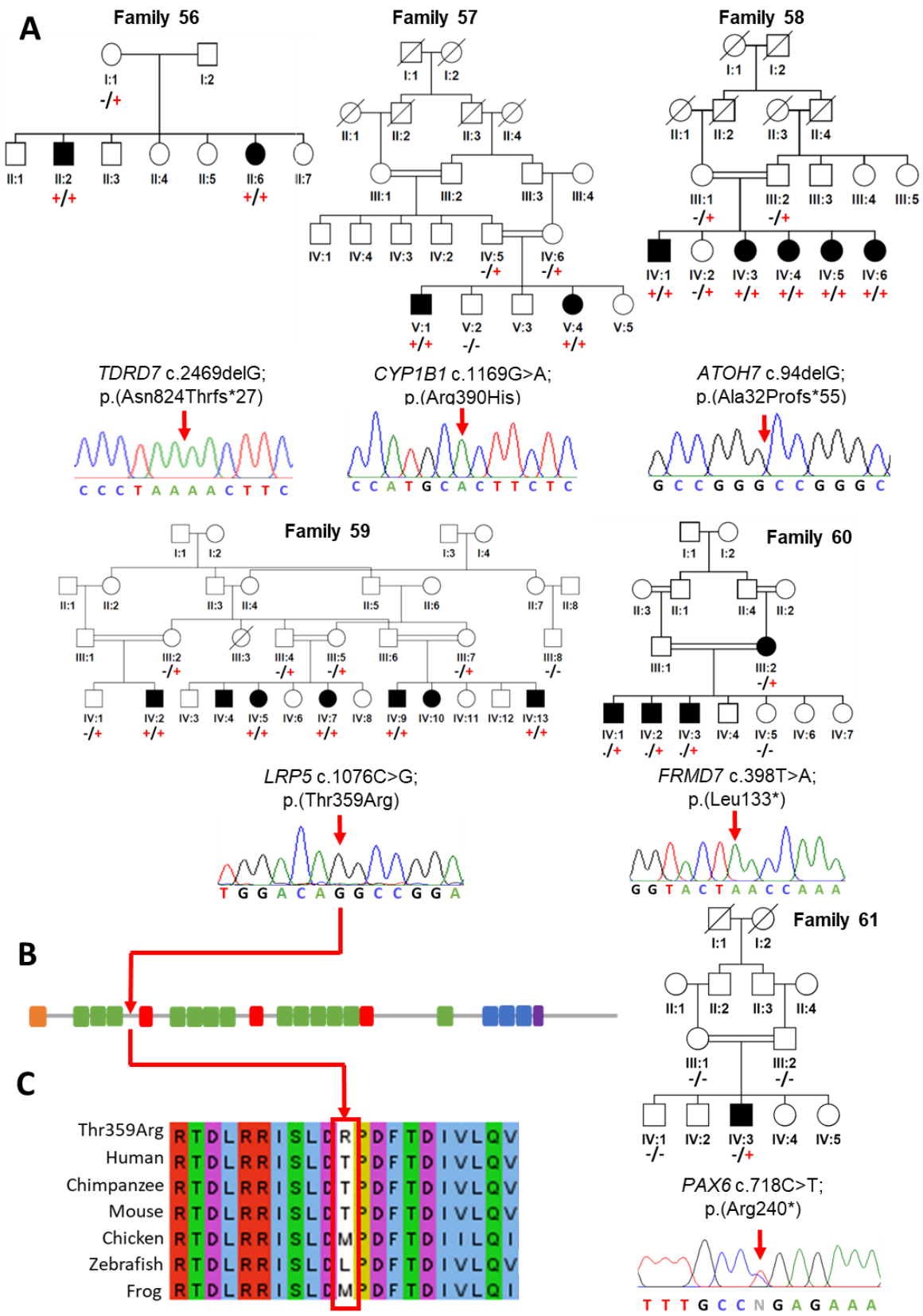

A
Family 56
Family 57
Family 58
TDRD7 c.2469delG;
p.(Asn824Thrfs*27)
CYP1B1 c.1169G>A;
p.(Arg390His)
ATOH7 c.94delG;
p.(Ala32Profs*55)
Family 59
Family 60
LRP5 c.1076C>G;
p.(Thr359Arg)
FRMD7 c.398T>A;
p.(Leu133*)
B
C
Family 61
Thr359Arg
Human
Chimpanzee
Mouse
Chicken
Zebrafish
Frog
PAX6 c.718C>T;
p.(Arg240*)

Figure 5.4 Pedigrees and genotype data for families 56 - 61

A) Pedigree diagrams of families 56 - 61 showing segregation of identified ocular disease *variants. Genotypes are shown beneath each family member investigated (+, variant; -, wild type) alongside sequence chromatograms of the* disease *variant in a homozygous (families 56-59), hemizygous (family 60) and heterozygous (family 61) affected individual in each family.*
(B) Schematic showing domain architecture of LRP5 [adapted from UniProt (216)] and the location of the c.1076C>G; p.(Thr359Arg) variant. Low-density lipoprotein receptor repeat class A and B domains are denoted by the blue and green rectangles respectively, orange rectangle represents the signal peptide, red rectangles denote the coagulation Factor Xa inhibitory site, and the purple rectangle denotes the transmembrane domain.
(C) Conservation analysis: multiple species alignments of partial amino acid sequences of LRP5 showing conservation of threonine at position 359.

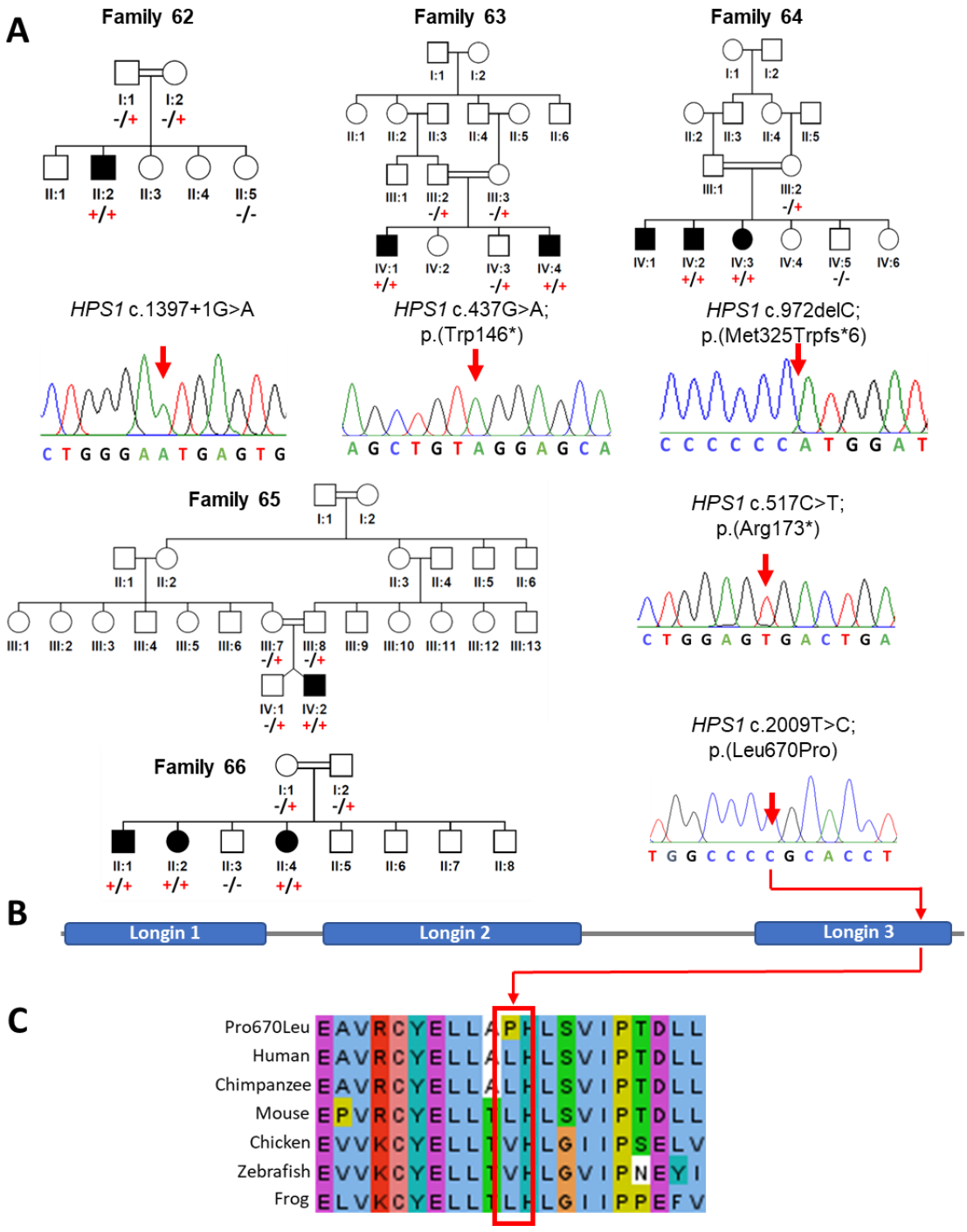


Figure 5.5 Pedigrees and *HPS1* genotype data for families 62 - 66

(A) Pedigree diagrams of families 62 - 66 showing segregation of identified HPS1 *variants, cosegregating appropriately for an autosomal recessive condition in each family investigated. Genotypes are shown beneath each family member investigated (+, variant; -, wild type) alongside sequence chromatograms of the* HPS1 *variant in a homozygous affected individual in each family.*
(B) Schematic showing domain architecture of HPS1 [adapted from UniProt (216)] and the location of the c.2009T>C; p.(Leu670Pro) variant within the Longin 3 domain.
(C) Conservation analysis: multiple species alignments of partial amino acid sequences of HPS1 showing conservation of leucine at position 670.

Table 5.3 Ocular findings in affected individuals homozygous for *FYCO1* c.2206C>T; p.(Gln736*)

Family (ID)	Age examined (yrs)	BCVA (Snellen)	IOP (mmHg)	Cornea	AC	Cataracts	Cataract surgery	Fundus
52 (V:3)	21	6/12	14	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* RE, decentered LE	RE: slight optic disc pallor, healthy retina LE: healthy optic disc & retina
52 (V:6)	13	HM	13	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ*	Healthy optic disc & retina
52 (V:6)	16	PL	15	Clear	D&Q	✓ BE present at birth RE lamellar cataract	✓ LE only, IOL *in situ*	No fundal view
52 (VI:3)	3	RE: 6/18 LE: 6/9	15	Clear	D&Q	✓ BE present at birth	✓ Aphakic	Healthy optic disc & retina
53 (V:1)	7	RE: 6/60 LE: 6/60	RE: 11 LE: 12	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* with PC thickening	Healthy optic disc, retina & macula
53 (V:2)	4	RE: 6/12 LE: 6/9	RE: 16 LE: 16	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* with PC thickening	Healthy optic disc & retina
54 (V:1)	14	RE: 6/12 LE: 6/12	10	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* with posterior capsulotomy	Healthy optic disc & retina
54 (V:3)	6	PL	8	Clear	D&Q	✓ BE present at birth Lamellar and posterior subcapsular cataract in un-operated eye	✓ Unilateral	No fundal view, dense vitritis on B-scan ultrasonography
55 (III:3)	26	RE: 6/9 LE: 6/12	9	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* with posterior capsulotomy	Healthy optic disc & retina
55 (III:4)	23	RE: 6/12 LE: 6/9	RE: 16 LE: 16	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ* with PC thickening	Healthy optic disc & retina

55 (III:5)	17	RE: 6/12 LE: 6/12	13	Clear	D&Q	✓ BE present at birth	✓ IOL *in situ*	Healthy optic disc & retina
55 (IV:3)	12	PL	RE: 10 LE: 10	Clear	D&Q	✓ BE present at birth Cortical cataract in un-operated eye	✓ Unilateral	No fundal view

Abbreviations: AC, anterior chamber; BCVA, best corrected visual acuity; BE, both eyes; D&Q, deep and quiet; HM, hand movement; ID, individual; IOL, intraocular lens; IOP, intraocular pressure; LE, left eye; NAD, no abnormalities detected; PC, posterior capsule; PL, perception of light; RE, right eye; yrs, years.

Table 5.4 Ocular findings in affected individuals homozygous for *ATOH7* c.94delG; p.(Ala32Profs*55)

Family (ID)	Age examined (yrs)	Age of onset (yrs)	Laterality	Progression	BCVA (Snellen)	IOP (mmHg)	Cornea	AC	Lens	Fundus	Ocular surgery
58 (1V:1)	18	Birth	Bilateral	Stable	HM	26	Cloudy, BK	NA	Cataract	No view due to cataract & BK	Nil
58 (1V:3)	15	Birth	Unilateral	Progressing	NPL	Hypotony	NA	Shallow, phthisical	Cataract	No view	Glaucoma surgery
58 (1V:4)	10	Birth	Unilateral	Stable	CF	6	Central stromal cornea opacity	Iridocorneal adhesions	IOL *in situ*	Detached retina with multiple adhesions	Cataract surgery
58 (1V:5)	8	Birth	Bilateral	Stable	BE: 3/60	7	BK	Shallow	IOL, decentred	Hypoplastic optic disc, retinal flat	Cataract surgery
58 (1V:6)	5	Birth	Bilateral	Progressing	BE: HM	18	Cloudy, megalocornea	Shallow, PI	Clear	Glaucomatous optic atrophy, retina flat	Glaucoma surgery

Abbreviations: AC, anterior chamber; BCVA, best corrected visual acuity; BE, both eyes; BK, band keratopathy; CF, counting fingers; HM, hand movement; IOL, intraocular lens; IOP, intraocular pressure; NA, no information available; NPL, no perception of light; PI, peripheral iridotomy; PL, perception of light; yrs, years

Table 5.5 Ocular findings in affected individuals homozygous for *LRP5* c.1076C>G; p.(Thr359Arg)

Family (ID)	Age examined (yrs)	Age of onset	Progression	Nystagmus	VA	Globe	Cornea	Iris adhesions	Cataract
59 (1V:2)	18	Early infancy	Stable	✓	Blind	Microphthalmia	Central opacity	✓	✓
59 (1V:5)	16	Early infancy	Stable	✓	Blind	Microphthalmia	Central opacity	✓	✓
59 (1V:7)	15	Early infancy	Stable	✓	Blind	Microphthalmia	Central opacity	✓	✓
59 (1V:9)	17	Early infancy	Stable	✓	Blind	Microphthalmia	Central opacity	✓	✓
59 (IV:13)	15	Early infancy	Stable	✓	Blind	Microphthalmia	Central opacity	✓	✓

Abbreviations: VA, visual acuity; yrs, years. (✓) denotes the presence of a feature

5.3.4 Discussion

Anophthalmia and microphthalmia are severe congenital developmental defects of the eye. In the clinical context, anophthalmia refers to a complete absence of the globe in the orbit, whilst microphthalmia refers to the presence of a small globe within the orbit. Both anophthalmia and microphthalmia are more commonly bilateral, although they can also present unilaterally. These are relatively rare defects, occurring with an estimated combined incidence of 1 in 10,000 live births (410). Anophthalmia and microphthalmia can occur as isolated malformations, or as part of a syndrome; both forms have been associated with autosomal recessive, autosomal dominant and X-linked patterns of inheritance, and display extensive genetic heterogeneity (411). *SOX2* variants are the major single-gene cause of anophthalmia and microphthalmia, accounting for ~10 - 15% of all cases (412). Variants in other genes have been shown to account for another ~25% of cases of anophthalmia and microphthalmia (413). In up to 50 - 60% of cases however, the underlying genetic cause remains undetermined (411, 414).

Fares-Taie *et al* first provided evidence in 2013 for *ALDH1A3* involvement in autosomal recessive anophthalmia and microphthalmia in humans (412). Since then, including the individuals reported in this study, *ALDH1A3* variants have been identified as a cause of autosomal recessive anophthalmia and microphthalmia in 67 individuals from 28 families to date (Table 5.6). Among these, 56 individuals are from 19 consanguineous families (412, 415-425) and even where consanguinity was not documented, most affected individuals were found to be homozygous for the disease-causing *ALDH1A3* variant, with only two reported individuals with anophthalma or microphthalmia associated with compound heterozygosity for pathogenic *ALDH1A3* variants (426, 427).

ALDH1A3-associated anophthalmia and microphthalmia is frequently reported in association with other ocular and extraocular anomalies, such as the presence of short eyelids, blepharophimosis and reduced palpebral fissures (420, 421, 425, 426), entropion (425), conjunctival symblepharon (421), conjunctival discoloration (421), large eyebrows and synophrys (421, 423), coloboma (416, 417, 419, 421, 425), hypoplasia of the optic tracts and chiasm (412, 415, 421, 423, 425), hypoplastic extra

ocular muscles (415, 423), refractive errors including both myopia and hyperopia (416, 417), and esotropia (416). There is a high variability observed in the phenotypic expression of dysmorphic or systemic features associated with anophthalmia and microphthalmia, even in individuals with the same *ALDH1A3* gene variants (418, 423, 424). Mild hypoplasia of the vermis (variant of Dandy-Walker malformation), as well as pulmonary stenosis and atrial septal defect, have previously been reported in association with *ALDH1A3*-associated anophthalmia and microphthalmia (412, 423). However, as these systemic findings have only been reported in a single individual each, it still remains unclear if these are associated or unrelated features of *ALDH1A3* variation. Occasionally, patients with *ALDH1A3*-associated anophthalmia and microphthalmia are also reported to have neurocognitive or behavioural features including intellectual disability, developmental delay and autism (412, 416, 417, 423). This association remains controversial due to the wide inter-familial variability in neurocognitive or behavioural outcomes (416, 417, 423), and the important confounding impact of visual impairment during development (428, 429). In addition, intellectual disability due to other genetic disorders may be more common in populations with high consanguinity (430).

It has previously been suggested that the difference in phenotype between microphthalmia and anophthalmia may be the result of residual ALDH1A3 activity (421). However, a review of all known disease-causing mutations in *ALDH1A3* (Figure 5.6 and Table 5.6) does not seem to support this hypothesis, with no consistent correlation between a particular phenotype (anophthalmia or microphthalmia) and the nature of variation (missense, nonsense, frameshift or splice variants) or the protein domain affected (NAD-binding domain, catalytic domain or oligomerisation domain). This may partly be due to difficulties in distinguishing between anophthalmia from severe microphthalmia in routine clinical practice. True congenital anophthalmia can only be diagnosed radiologically or histologically, and most published cases of clinical anophthalmia probably include cases of severe microphthalmia, where residual ocular tissue may have been present in the orbit despite external appearances of an absent globe (410). Although systemic features have been reported in *ALDH1A3*-associated ocular disease, these are uncommon, and the associations are controversial, providing a relatively good prognosis for affected families when compared to other known causes of anophthalmia.

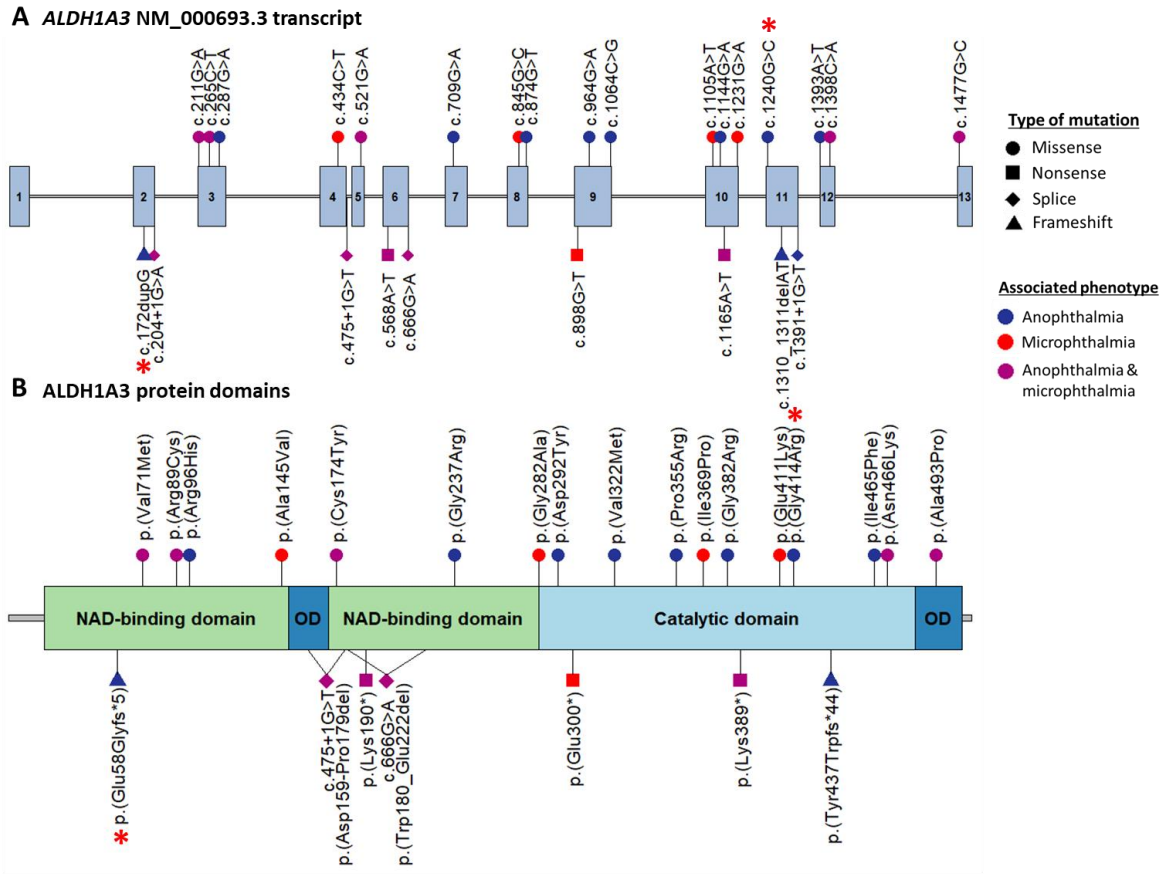


Figure 5.6 *ALDH1A3* variants associated with anopthalmia and microphthalmia.

(A) Schematic representation of ALDH1A3 NM_000693.3 transcript highlighting the position of all disease-causing variants associated with anophthalmia and microphthalmia identified to date.
(B) Schematic showing domain architecture of ALDH1A3 [predicted domains described by Moretti et al (24)] and the locations of all disease-causing variants associated with anophthalmia and microphthalmia identified to date. OD, oligomerisation domain.
Variants identified in this study are denoted by a red asterix (*).

To date, there have only been six previously reported variants in *TDRD7* associated with juvenile or congenital cataracts in homozygous form in six families (389, 431-434), including a further frameshift deletion [c.1129delG; p.(Ala377Profs*2)] in one other consanguineous Pakistani family (389) (Table 5.7). The identification of a second *TDRD7* variant in Pakistan as part of this study raises the possibility that *TDRD7* variation is a potentially under-recognised cause of congenital cataracts in Pakistan. Additionally, there have been a small number of paediatric cataract cases associated with monoallelic *TDRD7* variants (431, 435, 436), and there have also been reports of two families where homozygosity for *TDRD7* variants in male individuals was associated with azoospermia in combination with congenital cataracts (434); the true causal relationship in these cases however still remains uncertain.

Although affected individuals in families 57 and 58 were initially diagnosed with congenital cataracts, genomic studies subsequently identified homozygous sequence variants in *CYP1B1* and *ATOH7* respectively as likely responsible for ocular disease in these families. Autosomal recessive mutations in *CYP1B1* are the most common cause of primary congenital glaucoma, and the *CYP1B1* c.1169G>A; p.(Arg390His) variant has been found to be particularly common in Pakistani, North Indian, Iranian and Chinese communities (392, 437-439), supporting a revised clinical diagnosis of primary congenital glaucoma in this family. The initial misdiagnosis of congenital cataracts could have been due to difficulties in differentiating between leukocoria caused by lenticular opacities, and corneal oedema and opacity resulting from raised intraocular pressures in primary congenital glaucoma.

Disruption of *ATOH7* function has been previously described to cause a spectrum of developmental eye disorders with overlapping phenotypes in only a small number of families (Table 5.8). Associated conditions include non-syndromic congenital retinal nonattachment (24, 440, 441), persistent hyperplasia of the primary vitreous (442), and FEVR (443), with common clinical features including poor vision from birth, nystagmus, microphthalmia, corneal opacities, cataracts, vitreous anomalies including persistent foetal vasculature or a retrolental fibrovascular mass, retinal vascular abnormalities and retinal detachment. *ATOH7* variants have also been potentially implicated in optic nerve hypoplasia and aplasia (444, 445), although the true association remains unclear (442, 444, 446). Additionally, genome-wide association studies have implicated *ATOH7* variants in association with adult-onset open-angle glaucoma (447, 448), and a heterozygous *ATOH7* deletion has been associated with primary open angle glaucoma in a single individual (449); however the identification of two affected individuals (IV:1 and IV:4) in family 58 with features of congenital glaucoma (Table 5.2) represents the first time *ATOH7* variants have been associated with this phenotype. This is of potential clinical relevance, as in contrast to the permanent visual field loss associated with glaucomatous optic neuropathy in adults with adult-onset glaucoma, optic neuropathy in infants and young children with congenital glaucoma is potentially reversible, especially in the early stages of the disease (450). Detailed clinical phenotyping in additional families will be helpful to clarify this potential genotype-phenotype correlation, and a prompt ophthalmic assessment in all individuals with *ATOH7*-associated ocular disease may allow early

recognition and appropriate surgical management of congenital glaucoma if present, significantly improving a child's visual prognosis.

In family 59, identification of a homozygous *LRP5* variant causing global ocular developmental defects has permitted a refined clinical diagnosis and improved understanding of disease in affected individuals who did not have a previous ophthalmic diagnosis. In addition to ocular findings, *LRP5* variants are typically associated with reduced bone mineral density, osteopenia and osteoporosis (451, 452), and appropriate clinical screening of affected individuals may enable early detection of skeletal co-morbidities and may prevent the development of secondary complications.

Early onset nystagmus in an otherwise seemingly healthy child can be associated with a diverse range of ophthalmic conditions including achromatopsia, congenital stationary night blindness, inherited retinal dystrophies, aniridia and ocular albinism (144). In family 60, a novel *FRMD7* variant was identified as likely responsible for X-linked congenital idiopathic nystagmus in affected individuals. There are now over 90 *FRMD7* variants reported to cause congenital idiopathic nystagmus (453); this is however the first time *FRMD7* has been identified as a cause of congenital idiopathic nystagmus in Pakistani communities. The *FRMD7* molecular diagnosis in this family provided reassurance for the family regarding the good visual prognosis and non-progressive nature of the condition, without the need for complex, invasive and expensive investigations to rule out an underlying ocular disease. In family 61, despite the consanguineous family structure, identification of a well reported *PAX6* nonsense variant (6, 393-400) occurring *de novo* in the affected individual allowed accurate family counselling regarding recurrence risks. Additionally, this provided a molecular mechanism for the aniridia later identified on re-examination of the affected individual, and afforded reassurance by excluding a *PAX6* deletion extending to the *WT1* gene predisposing to Wilms tumour development (454).

HPS is an autosomal recessive condition where defects in intracellular vesicle formation and trafficking result in OCA associated with a bleeding diathesis and ceroid-lipofuscin lysosomal storage disease. This can result in life-threatening manifestations such as pulmonary fibrosis, granulomatous colitis, and immunodeficiency (405).

Awareness of this diagnosis and its associated systemic implications in families 62 - 66 has enabled counselling about the risk of pulmonary fibrosis, and importance of smoking cessation and prompt treatment of respiratory infections to optimise pulmonary function. Clinicians caring for the family are also now aware of the risk of prolonged bleeding and increased need for platelet or red blood cell transfusions following surgical procedures. The five disease-causing *HPS1* variants identified in this study also now doubles the number of *HPS1* variants reported in Pakistani families (455).

The families investigated in this study highlight the notable diagnostic difficulties encountered by clinicians in geographically isolated regions in Pakistan, who do not always have access to the specialised equipment for detailed ocular phenotyping necessary for an accurate ophthalmic diagnosis. Recognising these genes and variants as a cause of ocular disease within these communities will facilitate more accurate clinical diagnoses, early recognition and treatment for individuals with a similar clinical presentation in the region.

Table 5.6 Summary of all reported *ALDH1A3* variants associated with anophthalmia and microphthalmia

Type of variant	Nucleotide variant	Protein variant	Ethnicity	Number of affected families (individuals)	Associated phenotype	Reference	ClinVar (Accession)
Missense	c.211G>A	p.(Val71Met)	Israeli	1 (9)	Anophthalmia & microphthalmia	Mory *et al* (415)	Pathogenic (VCV000091908)
	c.265C>T	p.(Arg89Cys)	Pakistani	1 (2)	Anophthalmia & microphthalmia	Fares-Taie *et al* (412)	Pathogenic (VCV000040203)
	c.287G>A	p.(Arg96His)	Chinese	1 (2)	Anophthalmia	Liu *et al* (426)	Pathogenic (VCV000978214)
	c.434C>T	p.(Ala145Val)	Saudi Arabian	1 (2)	Microphthalmia	Aldahmesh *et al* (416)	Not present
	c.521G>A	p.(Cys174Tyr)	Lebanese	1 (3)	Anophthalmia & microphthalmia	Roos *et al* (417)	Not present
	c.709G>A	p.(Gly237Arg)	Chinese & Iranian	3 (5)	Anophthalmia	Liu *et al* (426); Dehghani *et al* (418)	Pathogenic (VCV000978215)
	c.845G>C	p.(Gly282Ala)	Arabic	1 (2)	Microphthalmia	Alabdullatif *et al* (419)	Not present
	c.874G>T	p.(Asp292Tyr)	Asian British - Bangladeshi	1 (2)	Anophthalmia	Patel *et al (427)*	Not present
	c.964G>A	p.(Val322Met)	Indian	1 (1)	Anophthalmia	Ullah *et al* (420)	Uncertain significance (VCV000653567)
	c.1064C>G	p.(Pro355Arg)	Egyptian	1 (1)	Anophthalmia	Abouzeid *et al* (421)	Not present
	c.1105A>T	p.(Ile369Pro)	Saudi Arabian	1 (3)	Microphthalmia	Aldahmesh *et al* (416)	Not present
	c.1144G>A	p.(Gly382Arg)	Egyptian	1 (4)	Anophthalmia	Abouzeid *et al* (421)	Not present
	c.1231G>A	p.(Glu411Lys)	Swiss - Sri Lankan	1 (1)	Microphthalmia	Abouzeid *et al* (421)	Not present

	c.1240G>C	p.(Gly414Arg)	Pakistani	1 (5)	Anophthalmia	Lin *et al* (422) – this study	Not present
	c.1393A>T	p.(Ile465Phe)	Asian British - Bangladeshi	1 (2)	Anophthalmia	Patel *et al (427)*	Not present
	c.1398C>A	p.(Asn466Lys)	Turkish	1 (2)	Anophthalmia & microphthalmia	Semerci *et al* (423)	Not present
	c.1477G>C	p.(Ala493Pro)	Turkish	1 (1)	Anophthalmia & microphthalmia	Fares-Taie *et al* (412)	Pathogenic (VCV000040204)
Nonsense	c.568A>T	p.(Lys190*)	Egyptian	1 (2)	Anophthalmia & microphthalmia	Yahyavi *et al* (425)	Not present
	c.898G>T	p.(Glu300*)	Swiss - Spanish	1 (1)	Microphthalmia	Abouzeid *et al* (421)	Not present
	c.1165A>T	p.(Lys389*)	Hispanic	1 (1)	Anophthalmia & microphthalmia	Yahyavi *et al* (425)	Not present
Splicing	c.204+1G>A	Alteration of WT donor site	Egyptian	1 (2)	Anophthalmia & microphthalmia	Abouzeid *et al* (421)	Not present
	c.475+1G>T	Skipping of exon 5	Moroccan	1 (1)	Anophthalmia & microphthalmia	Fares-Taie *et al* (412)	Pathogenic (VCV000040205)
	c.666G>A	Skipping of exon 6	Turkish	1 (7)	Anophthalmia & microphthalmia	Semerci *et al* (423); Plaisancié et al (424)	Not present
	c.1391+1G>T	Alteration of WT donor site	Egyptian	1 (1)	Anophthalmia	Abouzeid *et al* (421)	Not present
Frameshift	c.1310_1311 delAT	p.(Tyr437Trpfs*44)	Pakistani	1 (4)	Anophthalmia	Ullah *et al* (420)	Not present
	c.172dupG	p.(Glu58Glyfs*5)	Pakistani	1 (3)	Anophthalmia	Lin *et al* (422) – this study	Not present

Abbreviations: WT, wild type. Variants identified in this study are highlighted in yellow

Table 5.7 Summary of all reported *TDRD7* variants associated with inherited ocular disease

Type of variant	Mode of inheritance	Nucleotide variant	Protein variant	Ethnicity	Number of affected families (individuals)	Associated phenotype	Reference	ClinVar (Accession)
Missense	Monoallelic	c.83A>C	p.(Gln28Pro)	Han Chinese	1 (1)	Bilateral posterior polar cataract	Li *et al (435)*	Not present
	Biallelic (hom)	c.2539G>A	p.(Asp847Asn)	Saudi Arabian	1 (1)	AR cataract	Alfares *et al* (432)	Likely pathogenic (VCV000800981)
Nonsense	Biallelic (hom)	c.689dupA	p.(Tyr230*)	Chinese	1 (2)	AR congenital cataracts + azoospermia in males	Tan *et al* (434)	Pathogenic (VCV000427904)
	Monoallelic			Chinese	1 (1)	Cataract *(Also CRYBA2 variant het carrier)*	Sun *et al* (436)	
Frameshift	Biallelic (hom)	c.1129delG	p.(Ala377Profs*2)	Pakistani	1 (5)	AR congenital cataracts	Chen *et al* (389)	Pathogenic (VCV000426069)
	Biallelic (hom)	c.2469delG	p.(Asn824Thrfs*27)	Pakistani	1 (2)	AR congenital cataracts	This study	Not present
	Biallelic (hom)	c.328dupA	p.(Thr110Asnfs*30)	Chinese	1 (3)	AR congenital cataracts + azoospermia in males	Tan *et al* (434)	Pathogenic (VCV000427905)
Inframe deletion	Biallelic (hom)	c.1852_1854 delGGT	p.(Val618del)	NA	1 (4)	AR congenital cataracts	Lachke *et al* (431)	Not present
	Biallelic (hom)			NA	1 (1)	NA	Maddirevula *et al* (433)	
Chr 9 paracentric inversion disrupting *TDRD7 & NR5A1*	Monoallelic	46,X,Y,inv(9)(q22.33q34.11)		NA	1 (1)	Juvenile cataract and hypospadias *(NR5A1 disruption thought responsible for hypospadias)*	Lachke *et al* (431)	-

Abbreviations: AR, autosomal recessive; het, heterozygous; hom, homozygous. Variant identified in this study is highlighted in yellow.

Table 5.8 Summary of all reported biallelic *ATOH7* variants associated with inherited developmental ocular disease

Type of variant	Nucleotide variant	Protein variant	Ethnicity	Number of affected families (individuals)	Associated phenotype	Reference	ClinVar (Accession)
Missense	c.125G>C	p.(Arg42Pro)	Pakistani	1	Non-syndromic congenital retinal nonattachment	Keser *et al (440)*	Not present
	c.136A>C	p.(Asn46His)	Pakistani	1 (3)	Persistent hyperplasia of the primary vitreous	Prasov *et al* (442)	Pathogenic (VCV000144067)
	c.146A>T	p.(Glu49Val)	Pakistani	1 (5)	Ocular developmental defect: features include nystagmus, reduced vision, dense corneal opacity, microphthalmia, microcornea, retrolental mass	Khan *et al* (443)	Pathogenic (VCV000144065)
	c.175G>A	p.(Ala59Thr)	Swiss	1 (2)	Optic nerve hypoplasia, foveal hypoplasia and vascular abnormalities	Atac *et al* (444)	Pathogenic (VCV000812672)
	c.176C>T	p.(Ala59Val)	Swiss	1 (2)	Optic nerve hypoplasia, foveal hypoplasia and vascular abnormalities	Atac *et al* (444)	Uncertain significance (VCV000812673)
Frameshift	c.53delC	p.(Pro18Argfs*69)	Turkey	1 (2)	Familial exudative vitreoretinopathy	Khan *et al* (443)	Pathogenic (VCV000144066)
	c.94delG	p.(Ala32Profs*55)	Pakistani	1 (5)	Global ocular developmental defect: features include cataracts, congenital glaucoma, retinal detachment	This study	Not present

Inframe deletion	c.121_144del24	p.(Arg41_Arg48del)	Japanese	1(1)	Non-syndromic congenital retinal nonattachment	Kondo *et al* (24)	Uncertain significance (VCV001034468)
6523bp deletion in 5'UTR	c.-22208_-15686del6523	Deletion of transcriptional enhancer	Kurdish	1 (4)	Non-syndromic congenital retinal nonattachment	Ghiasvand *et al* (441)	Pathogenic (VCV000030807)
			Pakistani	2 (2)	Non-syndromic congenital retinal nonattachment	Keser *et al (440)*	

Abbreviations: hom, homozygous. Variant identified in this study is highlighted in yellow

5.4 Conclusions and future work

Studies of families with inherited ocular diseases in Pakistani and Palestinian communities have helped clarify the phenotypic characterisation and genetic associations of rare and ultra-rare inherited eye diseases, and provide much-needed scientific insight into the spectrum, nature and causes of ocular disease within these communities and globally.

Mitochondrial complex II deficiency is a rare inborn error of metabolism, accounting for approximately 2% of mitochondrial disease diagnoses. Only 61 patients have been described to date, associated with pathogenic variants in four complex II genes (*SDHA*, *SDHB*, *SDHD* and *SDHAF1*); of these, only two affected individuals have been reported with candidate biallelic variants in the *SDHD* gene (456). The three additional affected individuals described in this study further consolidate recessive *SDHD* variants as contributing to mitochondrial complex II deficiency. There is unfortunately no cure for complex II deficiency, although some therapeutic interventions may be associated with symptomatic improvements in these patients, including riboflavin, l-carnitine and ubiquinone (456). Further studies integrating detailed clinical phenotyping and metabolic investigations, molecular genomic analyses combined with *in silico* pathogenicity prediction and structural modelling, histological and histochemical studies of muscle biopsy specimens, spectrophotometric assays of OXPHOS respiratory chain function, BN-PAGE and SDS-PAGE followed by Western blot analysis to study physiological mitochondrial complex protein-protein interactions, and complementation studies in patient and yeast cells will be useful in developing a better understanding of the molecular basis of these conditions, which may lead to the development of more effective therapies.

TDRD7 (389, 431-434) and *ATOH7* (24, 440-443) are uncommon causes of developmental eye disorders. *ATOH7* dysfunction is associated with a wide spectrum of developmental ocular phenotypes including non-syndromic congenital retinal nonattachment (24, 440, 441), persistent hyperplasia of the

primary vitreous (442), and FEVR (443), and potentially implicated in optic nerve hypoplasia and aplasia (444, 445). *ATOH7* is known to function as an early retinal transcription factor crucial for RGC development and specification (441), and further functional characterisation in animal models or iPSC-derived retinal organoids may be useful in assessing the impact of pathogenic *ATOH7* variants on RGC morphology, and delineate the mechanisms by which *ATOH7* dysfunction leads to the diverse developmental ocular phenotypes described.

TDRD7 is a component of ribonucleoprotein complexes and RNA granules in differentiating lens cells, suggesting that post-transcriptional regulation of gene expression plays an important role in lens homeostasis (431). *TDRD* gene polymorphisms have potentially been associated with susceptibility to age-related cataracts (457), and there is also increasing evidence for additional genes implicated in the development of both congenital and adult-onset cataracts, such as *LIM2*, *EPHA2*, and *CRYAA* (458, 459). Understanding the genetic basis and molecular pathogenesis of congenital cataracts may lead to the development of strategies that prevent or delay the development of age-related cataracts, hence reducing the enormous global burden for cataract surgery. Studies of RNA and protein-binding targets of *TDRD7*, coupled with bioinformatics strategies, have identified several potential interacting partners of *TDRD7*, including *HSPB1*, *NRAS* and *ACTN2* (460-462). Further investigation into these molecular pathways, including subcellular localisation studies in lens fiber and epithelial cells, characterising gene expression in the lens during development, and studying ocular and lens development in xenopus, zebrafish or mouse gene-perturbation models (463-465), may yield new insights into cataractogenesis.

There is a wide phenotypic variation in *ALDH1A3*-associated ocular disease. Individuals with the same *ALDH1A3* variant can display both anophthalmia and microphthalmia in different eyes (419, 421, 425), and affected individuals with the same variant within the same family have been found to have clinical phenotypes of differing severity (417, 421, 423-425). Epidemiological studies have predicted the contribution of both genetic and environmental factors in the pathogenesis of congenital eye defects including anophthalmia and

microphthalmia (466), and the wide phenotypic spectrum seen may result from the impact of modifying genes or environmental influences affecting the *ALDH1A3*-associated eye disease phenotype. Further studies including DNA-methylation studies in animal models investigating epigenomic influences, transcriptomic analyses in patients linking altered gene expression to differential phenotypic outcomes, and co-immunoprecipitation and pull-down assays to investigate ALDH1A3 protein interactions, would be useful to define this interaction and elucidate underlying pathways. Given that *ALDH1A3* gene mutations appear to be the most common cause of anophthalmia and microphthalmia in consanguineous families (412, 415-425), screening for variants in this gene before exome analysis in populations with high rates of consanguinity can be considered.

6 CONCLUDING COMMENTS

This thesis documents the results of clinical and genetic analyses in a total of 66 Amish, Pakistani and Palestinian families, with a range of genetically undiagnosed syndromic and non-syndromic inherited ocular diseases. The work described in this thesis entails the identification of 18 novel pathogenic variants in 10 genes associated with a variety of inherited ocular disorders, including OCA (*TYR*, *OCA2*, *HPS1*), congenital nystagmus (*FRMD7*), ocular developmental disorders (*ALDH1A3*, *TDRD7*, *ATOH7* and *LRP5*) and ciliopathies (*BBS5*, *SCAPER*). This knowledge significantly expands on the molecular spectrum of inherited eye diseases in communities and globally. In many families investigated, the molecular diagnosis enabled a subsequent confirmation or refinement of the initial clinical diagnosis, as detailed in chapter 5.3, with important treatment and prognostic implications for affected individuals and their families.

In the Amish, Pakistani and Palestinian communities, geographical isolation, common ancestry and endogamy, combined with the often large family sizes typical of families in these regions, often results in an enrichment of disease-causing founder variants. These often represent important causes of disease in a particular region, and identification of the *INPP5E* c.1879C>T; p.(Gln627*) founder variant associated with MORM syndrome in the Punjab region in Pakistan (described in chapter 4.2), as well as the *SCAPER* c.2236dupA, p.(Ile746Asnfs*6) founder variant in the Ohio Amish community (described in chapter 4.4), allows targeted genetic testing for these variants, permitting a more rapid and cost-effective means of achieving a molecular diagnosis, shortening the diagnostic odyssey for affected individuals within the community.

There is a relative lack of knowledge regarding the specific nature and causes of inherited ocular diseases particularly in developing nations such as Pakistan. Literature reviews were conducted, compiling the most comprehensive and curated reports of genes and variants associated with BBS and OCA in Pakistan to date (presented in Appendices B and C respectively). This work

highlights the *OCA2* c.1045-15T>G and p.(Asp486Tyr) founder variants, as well as the *TYR* p.(Arg278*) recurrent variant as contributing significantly to OCA in Pakistan. These reviews will enhance knowledge of the genomic architecture pertaining to these inherited eye diseases in Pakistan, and facilitate the development of community-relevant disease databases correlating allele frequencies with geographical localisation and ethnicity. This is of enormous value for local healthcare resource planning, enabling the design of community-specific hierarchical strategies for rapid and cost-effective genetic testing arrays to enable an accurate disease diagnosis to be achieved more rapidly, thus aiding the development of diagnostic and clinical care pathways and policies throughout Pakistan.

This thesis also details comprehensive clinical and genetic studies of a number of rare inherited ocular disorders, described in only small numbers of affected individuals to date. This includes MORM syndrome and SCAPER syndrome (described in chapters 4.2 and 4.4), mitochondrial complex II deficiency associated with biallelic *SDHD* variation (chapter 5.2), *TDRD7*-associated congenital cataracts (chapter 5.3), and developmental eye defects associated with *ALDH1A3* and *ATOH7* (chapter 5.3). The deep phenotypic characterisation performed in these study families allowed a more comprehensive clinical comparison to be made with the previously reported cases, expanding on the disease-associated clinical spectrum, delineating the core and variable phenotypic features, and clarifying the presence (or indeed absence) of any genotype-phenotype correlation associated with variants in these genes.

The work of this thesis also provides novel insights into known inherited eye disease genes. The role of two common *TYR* variants p.(Ser192Tyr) and p.(Arg402Gln) in OCA, previously considered to represent non-pathogenic polymorphic alleles although subsequently contested, has been debated over the last 10 years. Chapter 3.2 now provides strong evidence that the *TYR* p.(Ser192Tyr)/p.(Arg402Gln) in *cis* haplotype may cause disease both when occurring in compound heterozygous fashion with a second deleterious allele, or in doubly homozygous state. The increased frequency of the *BBS1* p.(Met390Arg) variant in the Pennsylvania Amish community will similarly

permit empowered genetic and co-segregation studies able to determine phenotypic penetrance of this hypomorphic variant, which has important diagnostic implications given that this common founder variant is responsible for ~80% of BBS1 cases in European populations (297). Alongside detailed phenotypic assessments, this will allow clarity into the potential contribution of any additional BBS modifier variants, and may provide evidence supporting or disputing the controversial triallelic hypothesis in BBS. To this end, recruitment of additional family members in the extended Amish family with BBS studied in chapter 4.3 is ongoing, and detailed phenotype information is being obtained in collaboration with local ophthalmologists and clinicians.

Within this thesis, several sequencing techniques were employed, often in a hierarchical fashion, in the genomic investigation of families with inherited eye diseases. For instance, albinism families often first underwent *TYR* gene screening, followed by Illumina TruSight™ One clinical exome sequencing, in order to detect causative variants in known albinism genes, whilst consanguineous families with complex ocular phenotypes were instead investigated with genome-wide SNP genotyping and WES in the first instance, which would facilitate the detection of rare or even novel disease genes involved in hereditary eye diseases. Each genomic technique has recognised limitations and advantages, balancing a more comprehensive analysis against the difficulties in interpreting large numbers of variants; and the tailored approach adopted in this study enabled a pragmatic, efficient and cost-effective means of achieving a molecular diagnosis in many affected families, reflecting the likely real-word application of genomic technologies in these resource-scarce communities.

Studies in this thesis have contributed to a recent development of a new microarray platform diagnostic testing panel, led by the WoH project and now available in the Wisconsin Clinical Laboratory Improvement Amendments (CLIA) - certified regional testing laboratory. This flexible platform offers a rapid and cost-effective means of targeting the specific gene mutations of >150 genetic disorders identified across the Amish communities, and can be readily expanded to incorporate new disease-associated variants as they are identified

within the community. Testing via the platform is offered at cost to the Amish community, and has particular utility as part of a molecular-based newborn screening program, allowing affected individuals and families to benefit from early diagnosis and intervention. One particular benefit will be in the early diagnosis of propionic acidemia, an inborn error of metabolism occurring at an increased frequency in the Amish-Mennonite communities due to a founder missense variant in *PCCB* [c.1606A>G; p.(Asn536Asp)] (467). This is a condition where early diagnosis and intervention is shown to improve long-term neurological outcomes (468), however, cases may be missed using standard newborn screening metabolic assays. This genetic testing platform may also be of use as a carrier screening assay as the communities grow more receptive towards this, permitting anticipatory health care to prevent or mitigate the devastating effects of genetic diseases that may occur in the newborn.

In recent years, advances in NGS technologies and bioinformatics processing algorithms, tools, and pipelines, have led to improvements in the diagnosis of Mendelian disorders. Despite this, overall diagnostic rates for exome sequencing, the primary genomic technique employed in this study, is limited to ~25 - 50% (433). Some of these molecularly undiagnosed individuals may harbour variants in genes yet to be associated with human disease. Studies of inherited eye diseases in genetically isolated populations such as the Amish communities of the USA, or rural Pakistani and Palestinian communities, where the unique genomic architecture of these communities enables the application of empowered genomic approaches for disease gene identification (60, 469, 470), together with genomic matchmaking initiatives (471), will facilitate the identification of novel gene-disease associations, whilst also providing desperately required healthcare benefits for the families and populations involved.

The relatively low diagnostic rate of exome sequencing may also reflect the restricted focus, limited to coding sequences and intronic sequences flanking the exon-intron boundaries, representing only ~1.5% of the human genome. The application of whole genome sequencing approaches, which enable the detection of structural variants, deep intronic and regulatory variants in non-

coding regions, may lead to further improvements in diagnostic yield, and is an option being considered for any remaining molecularly undiagnosed individuals recruited through the established collaborative research consortium involving the community genomics research group at the University of Exeter, and collaborating scientists and clinicians within the Amish communities and at research centres in Pakistan and Palestine.

Interestingly, it has been suggested that genome sequencing only offers a minimal uplift in diagnostic yield compared to exome sequencing, and that the primary limitation to molecular diagnostic rates may relate not to the capture or calling of causal variants at the sequencing stage, but rather in the interpretation of the genomic variants identified (433). Current strategies for variant interpretation include family segregation, computational variant effect prediction, and functional assays (472), and are fairly reliable when applied to protein-coding regions of the genome, although predicting the pathogenicity of missense variants can be challenging (473), as demonstrated by the uncertain significance of the novel *OCA2* p.(Arg588Trp) variant identified in family 38 (chapter 3.3). Other sequence variants, such as deep intronic, enhancer or promoter region variants, as well as coding sequence variants with splice-disrupting potential, remain difficult to interpret, despite their clear contribution to Mendelian disease. Whole transcriptome sequencing approaches (RNA-seq) allowing quantification of gene expression levels and detection of alternative splicing (474), as well as multiplexed functional assays for variant effect (475), show promising potential in empowering variant interpretation and resolving the variants of uncertain significance “trapped in the interpretative void between benign and pathogenic” (476). These approaches are currently largely restricted to research environments, but their validation and integration into mainstream clinical practice may bring about transformative benefits in rare disease diagnostics.

Together, the work presented in this thesis provides an improved knowledge of the clinical and molecular spectrum of hereditary eye diseases in the Amish, Pakistan and Palestinian communities. These findings enabled targeted changes in clinical management and accurate prognostic information to be

provided to affected individuals and families, with communities benefitting from targeted genetic testing strategies that permit cost-effective disease diagnosis and improved carrier detection. This new knowledge may also enable experimental approaches for restoration of protein function through gene therapy, such as Luxturna® (voretigene neparvovec), the first ever licensed gene replacement therapy for the treatment of a genetic disease in humans, approved for the treatment of patients with *RPE65*-associated retinal dystrophy (12). In this way, an improved understanding of disease pathways arising from the discovery of novel disease genes identified in a community setting may thus ultimately also be applied to genetically diverse populations such as the UK and bring about clinical benefits for patients worldwide.

REFERENCES

1. Koeppen BM, Stanton BA. Berne and levy physiology e-book: Elsevier Health Sciences; 2017.

2. Jean D, Ewan K, Gruss P. Molecular regulators involved in vertebrate eye development. Mech Dev. 1998;76(1-2):3-18.

3. Shaham O, Menuchin Y, Farhy C, Ashery-Padan R. Pax6: a multi-level regulator of ocular development. Prog Retin Eye Res. 2012;31(5):351-76.

4. Walther C, Gruss P. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development. 1991;113(4):1435-49.

5. Puschel AW, Gruss P, Westerfield M. Sequence and expression pattern of pax-6 are highly conserved between zebrafish and mice. Development. 1992;114(3):643-51.

6. Glaser T, Walton DS, Maas RL. Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. Nat Genet. 1992;2(3):232-9.

7. Hill RE, Favor J, Hogan BL, Ton CC, Saunders GF, Hanson IM, et al. Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature. 1991;354(6354):522-5.

8. Plaisancie J, Ceroni F, Holt R, Zazo Seco C, Calvas P, Chassaing N, et al. Genetics of anophthalmia and microphthalmia. Part 1: Non-syndromic anophthalmia/microphthalmia. Hum Genet. 2019;138(8-9):799-830.

9. Snider TN, Department for B, Materials Sciences SoDUoMM, Mishina Y. Cranial neural crest cell contribution to craniofacial formation, pathology, and future directions in tissue engineering. Birth Defects research Part C, Embryo Today : Reviews. 2020;102(3):324-32.

10. Cavodeassi F, Creuzet S, Etchevers HC. The hedgehog pathway and ocular developmental anomalies. Hum Genet. 2019;138(8-9):917-36.

11. Lamb TD. Evolution of phototransduction, vertebrate photoreceptors and retina. Progress in retinal and eye research. 2013;36:52-119.

12. Graw J. The genetic and molecular basis of congenital eye defects. Nat Rev Genet. 2003;4(11):876-88.

13. Erskine L, Herrera E. Visual system development in vertebrates. eLS. 2013.

14. Cvekl A, Tamm ER. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. Bioessays. 2004;26(4):374-86.

15. Lim Z, Rubab S, Chan YH, Levin AV. Pediatric cataract: the Toronto experience-etiology. Am J Ophthalmol. 2010;149(6):887-92.

16. Hejtmancik JF. Congenital cataracts and their molecular genetics. Semin Cell Dev Biol. 2008;19(2):134-49.

17. Bell SJ, Oluonye N, Harding P, Moosajee M. Congenital cataract: a guide to genetic and clinical management. Therapeutic Advances in Rare Disease. 2020;1:2633004020938061.

18. Williams AL, Bohnsack BL. Neural crest derivatives in ocular development: discerning the eye of the storm. Birth Defects Res C Embryo Today. 2015;105(2):87-95.

19. Berry FB, Lines MA, Oas JM, Footz T, Underhill DA, Gage PJ, et al. Functional interactions between FOXC1 and PITX2 underlie the sensitivity to FOXC1 gene dose in Axenfeld-Rieger syndrome and anterior segment dysgenesis. Hum Mol Genet. 2006;15(6):905-19.

20. Tumer Z, Bach-Holm D. Axenfeld-Rieger syndrome and spectrum of PITX2 and FOXC1 mutations. Eur J Hum Genet. 2009;17(12):1527-39.

21. Ghazi NG, Green WR. Pathology and pathogenesis of retinal detachment. Eye (Lond). 2002;16(4):411-21.

22. Nissenkorn I, Department of Ophthalmology BMCPTI, Kremer I, Cohen S, Ben-Sira I. Nasal versus temporal preretinal vasoproliferation in retinopathy of prematurity. The British Journal of Ophthalmology. 1989;73(9):747-9.

23. Xiao H, Tong Y, Zhu Y, Peng M. Familial Exudative Vitreoretinopathy-Related Disease-Causing Genes and Norrin/β-Catenin Signal Pathway: Structure, Function, and Mutation Spectrums. Journal of ophthalmology. 2019;2019.

24. Kondo H, Matsushita I, Tahira T, Uchio E, Kusaka S. Mutations in ATOH7 gene in patients with nonsyndromic congenital retinal nonattachment and familial exudative vitreoretinopathy. Ophthalmic Genet. 2016;37(4):462-4.

25. Cepko CL, Austin CP, Yang X, Alexiades M, Ezzeddine D. Cell fate determination in the vertebrate retina. Proc Natl Acad Sci U S A. 1996;93(2):589-95.

26. Markitantova Y, Simirskii V. Inherited eye diseases with retinal manifestations through the eyes of homeobox genes. International journal of molecular sciences. 2020;21(5):1602.

27. Sharon D, Sandberg MA, Caruso RC, Berson EL, Dryja TP. Shared mutations in NR2E3 in enhanced S-cone syndrome, Goldmann-Favre syndrome, and many cases of clumped pigmentary retinal degeneration. Archives of ophthalmology. 2003;121(9):1316-23.

28. Demirbilek H, Hatipoglu N, Gul U, Tatli ZU, Ellard S, Flanagan SE, et al. Permanent neonatal diabetes mellitus and neurological abnormalities due to a novel homozygous missense mutation in NEUROD1. Pediatric diabetes. 2018;19(5):898-904.

29. Hull S, Arno G, Plagnol V, Chamney S, Russell-Eggitt I, Thompson D, et al. The phenotypic variability of retinal dystrophies associated with mutations in CRX, with report of a novel macular dystrophy phenotype. Invest Ophthalmol Vis Sci. 2014;55(10):6934-44.

30. Bessant DA, Holder GE, Fitzke FW, Payne AM, Bhattacharya SS, Bird AC. Phenotype of retinitis pigmentosa associated with the Ser50Thr mutation in the NRL gene. Arch Ophthalmol. 2003;121(6):793-802.

31. Hendrickson AE, Yuodelis C. The morphological development of the human fovea. Ophthalmology. 1984;91(6):603-12.

32. Kondo H. Foveal hypoplasia and optical coherence tomographic imaging. Taiwan journal of ophthalmology. 2018;8(4):181.

33. Gregory-Evans CY, Gregory-Evans K. Foveal hypoplasia: the case for arrested development. Expert Review of Ophthalmology. 2011;6(5):565-74.

34. Gregory-Evans CY, Wallace VA, Gregory-Evans K. Gene networks: dissecting pathways in retinal development and disease. Progress in retinal and eye research. 2013;33:40-66.

35. Sang DN. Embryology of the vitreous. Congenital and developmental abnormalities. Bull Soc Belge Ophtalmol. 1987;223 Pt 1:11-35.

36. Le Goff M, Bishop P. Adult vitreous structure and postnatal changes. Eye. 2008;22(10):1214-22.

37. Sebag J. Vitreous: in health and disease: Springer; 2014.

38. Muggenthaler MM, Chowdhury B, Hasan SN, Cross HE, Mark B, Harlalka GV, et al. Mutations in HYAL2, Encoding Hyaluronidase 2, Cause a Syndrome of Orofacial Clefting and Cor Triatriatum Sinister in Humans and Mice. PLoS Genet. 2017;13(1):e1006470.

39. Ten Kate LP, Al-Gazali L, Anand S, Bittles A, Cassiman JJ, Christianson A, et al. Community genetics. Its definition 2010. J Community Genet. 2010;1(1):19-22.

40. Amish Population, 2020 Elizabethtown College: Young Center for Anabaptist and Pietist Studies; 2020 [Available from: http://groups.etown.edu/amishstudies/statistics/statistics-population-2020/.

41. Kraybill DB, Johnson-Weiner KM, Nolt SM. The Amish: JHU Press; 2013.

42. Patel H, Cross H, Proukakis C, Hershberger R, Bork P, Ciccarelli FD, et al. SPG20 is mutated in Troyer syndrome, an hereditary spastic paraplegia. Nat Genet. 2002;31(4):347-8.

43. McKusick VA. Ellis-van Creveld syndrome and the Amish. Nat Genet. 2000;24(3):203-4.

44. Alkuraya FS. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. Hum Genet. 2013;132(11):1197-211.

45. Arcos-Burgos M, Muenke M. Genetics of population isolates. Clin Genet. 2002;61(4):233-47.

46. McKusick VA, Hostetler JA, Egeland JA. Genetic studies of the Amish, background and potentialities. Bull Johns Hopkins Hosp. 1964;115:203-22.

47. Francomano CA, McKusick VA, Biesecker LG. Medical genetic studies in the Amish: historical perspective. Am J Med Genet C Semin Med Genet. 2003;121c(1):1-4.

48. Cross HE, Crosby AH. Amish contributions to medical genetics. Mennonite Quarterly Review. 2008;82(3):449-68.

49. McKusick VA. Medical genetic studies of the Amish: selected papers: Johns Hopkins University Press; 1978.

50. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet. 2007;80(4):588-604.

51. McKusick-Nathans Institute of Genetic Medicine. Online Mendelian Inheritance in Man (OMIM) Baltimore, MD [Available from: https://omim.org/.

52. Payne M, Rupar CA, Siu GM, Siu VM. Amish, mennonite, and hutterite genetic disorder database. Paediatr Child Health. 2011;16(3):e23-4.

53. DESA U. World Population Prospects 2019. United Nations. Department of Economic and Social Affairs. World Population Prospects 2019. 2019.

54. Hussain R. The effect of religious, cultural and social identity on population genetic structure among Muslims in Pakistan. Ann Hum Biol. 2005;32(2):145-53.

55. Pigeyre M, Saqlain M, Turcotte M, Raja GK, Meyre D. Obesity genetics: insights from the Pakistani population. Obes Rev. 2018;19(3):364-80.

56. Bittles AH, Small NA. Consanguinity, genetics and definitions of kinship in the UK Pakistani population J Biosoc Sci. 2016;48(6):844-54.

57. Qasim I, Ahmad B, Khan MA, Khan N, Muhammad N, Basit S, et al. Pakistan Genetic Mutation Database (PGMD); A centralized Pakistani mutome data source. Eur J Med Genet. 2018;61(4):204-8.

58. Merten M. Keeping it in the family: consanguineous marriage and genetic disorders, from Islamabad to Bradford. Bmj. 2019;365:l1851.

59. Riaz M, Tiller J, Ajmal M, Azam M, Qamar R, Lacaze P. Implementation of public health genomics in Pakistan. Eur J Hum Genet. 2019;27(10):1485-92.

60. Li L, Jiao X, D'Atri I, Ono F, Nelson R, Chan CC, et al. Mutation in the intracellular chloride channel CLCC1 associated with autosomal recessive retinitis pigmentosa. PLoS Genet. 2018;14(8):e1007504.

61. Khan S, Lin S, Harlalka GV, Ullah A, Shah K, Khalid S, et al. BBS5 and INPP5E mutations associated with ciliopathy disorders in families from Pakistan. Ann Hum Genet. 2019;83(6):477-82.

62. Palestinian Central Bureau of Statistics. Palestine in figures 2019. Ramallah, Palestine; 2020.

63. Zlotogora J. Molecular basis of autosomal recessive diseases among the Palestinian Arabs. Am J Med Genet. 2002;109(3):176-82.

64. Abu-Libdeh B, Turnpenny PD, Teebi A. Genetic Disease in Palestine and Palestinians. In: Kumar D, editor. Genomics and health in the developing world: Oxford University Press; 2012.

65. Jaber L, Bailey-Wilson JE, Haj-Yehia M, Hernandez J, Shohat M. Consanguineous matings in an Israeli-Arab community. Arch Pediatr Adolesc Med. 1994;148(4):412-5.

66. Schellekens J, Kenan G, Hleihel A. The decline in consanguineous marriage among Muslims in Israel: The role of education. Demographic Research. 2017;37:1933-48.

67. Sirdah MM. Consanguinity profile in the Gaza Strip of Palestine: large-scale community-based study. Eur J Med Genet. 2014;57(2-3):90-4.

68. Na'amnih W, Romano-Zelekha O, Kabaha A, Rubin LP, Bilenko N, Jaber L, et al. Continuous decrease of consanguineous marriages among Arabs in Israel. Am J Hum Biol. 2015;27(1):94-8.

69. Zlotogora J, Barges S, Bisharat B, Shalev SA. Genetic disorders among Palestinian Arabs. 4: Genetic clinics in the community. Am J Med Genet A. 2006;140(15):1644-6.

70. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. Br J Ophthalmol. 2012;96(5):614-8.

71. Dandona L, Dandona R. Revision of visual impairment definitions in the International Statistical Classification of Diseases. BMC Med. 2006;4:7.

72. Pizzarello L, Abiose A, Ffytche T, Duerksen R, Thulasiraj R, Taylor H, et al. VISION 2020: The Right to Sight: a global initiative to eliminate avoidable blindness. Arch Ophthalmol. 2004;122(4):615-20.

73. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020--the right to sight. Bull World Health Organ. 2001;79(3):227-32.

74. Moore T, Burton H, Alberg C. Genetic ophthalmology in focus. A needs assessment and review of specialist services for genetic eye disorders Foundation for Genomics and Population Health. 2008.

75. Bunce C, Zekite A, Wormald R, Bowman R. Is there evidence that the yearly numbers of children newly certified with sight impairment in England and Wales has increased between 1999/2000 and 2014/2015? A cross-sectional study. BMJ Open. 2017;7(9):e016888.

76. Rahi JS, Cable N. Severe visual impairment and blindness in children in the UK. Lancet. 2003;362(9393):1359-65.

77. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. BMJ Open. 2014;4(2):e004015.

78. Quartilho A, Simkiss P, Zekite A, Xing W, Wormald R, Bunce C. Leading causes of certifiable visual loss in England and Wales during the year ending 31 March 2013. Eye (Lond). 2016;30(4):602-7.

79. Fincham GS, Pasea L, Carroll C, McNinch AM, Poulson AV, Richards AJ, et al. Prevention of retinal detachment in Stickler syndrome: the Cambridge prophylactic cryotherapy protocol. Ophthalmology. 2014;121(8):1588-97.

80. Ansley SJ, Badano JL, Blacque OE, Hill J, Hoskins BE, Leitch CC, et al. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature. 2003;425(6958):628-33.

81. Weiss JS, Moller HU, Lisch W, Kinoshita S, Aldave AJ, Belin MW, et al. The IC3D classification of the corneal dystrophies. Cornea. 2008;27 Suppl 2:S1-83.

82. Weiss JS, Moller HU, Aldave AJ, Seitz B, Bredrup C, Kivela T, et al. IC3D classification of corneal dystrophies--edition 2. Cornea. 2015;34(2):117-59.

83. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. Prog Retin Eye Res. 2010;29(5):335-75.

84. Gupta S, Chatterjee S, Mukherjee A, Mutsuddi M. Whole exome sequencing: Uncovering causal genetic variants for ocular diseases. Exp Eye Res. 2017;164:139-50.

85. Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, et al. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. Am J Hum Genet. 2017;100(1):75-90.

86. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, et al. Integrating Genomics into Healthcare: A Global Responsibility. Am J Hum Genet. 2019;104(1):13-20.

87. Jaeger W. Horner's law. The first step in the history of the understanding of X-linked disorders. Ophthalmic Paediatr Genet. 1992;13(2):49-56.

88. Moriwaki YI, D. A probable case of cytoplasmic inheritance in man: A critique of leber's disease. Journal of Genetics. 2020;33(2):163-7.

89. Jaeger C, Jay B. X-linked ocular albinism: a family containing a manifesting heterozygote, and an affected male married to a female with autosomal recessive ocular albinism. Hum Genet. 1981;56(3):299-304.

90. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68(4):820-3.

91. Kajiwara K, Berson EL, Dryja TP. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science. 1994;264(5165):1604-8.

92. Wright AF, Chakarova CF, Abd El-Aziz MM, Bhattacharya SS. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nat Rev Genet. 2010;11(4):273-84.

93. DiCarlo JE, Mahajan VB, Tsang SH. Gene therapy and genome surgery in the retina. J Clin Invest. 2018;128(6):2177-88.

94. Ziccardi L, Cordeddu V, Gaddini L, Matteucci A, Parravano M, Malchiodi-Albedi F, et al. Gene Therapy in Retinal Dystrophies. Int J Mol Sci. 2019;20(22).

95. National Institute for Health and Care Excellence. Voretigene neparvovec for treating inherited retinal dystrophies caused by RPE65 gene mutations. NICE Highly Specialised Technologies Guidance [HST11]. 2019.

96. Ramlogan-Steel CA, Murali A, Andrzejewski S, Dhungel B, Steel JC, Layton CJ. Gene therapy and the adeno-associated virus in the treatment of genetic and acquired ophthalmic diseases in humans: Trials, future directions and safety considerations. Clin Exp Ophthalmol. 2019;47(4):521-36.

97. Cideciyan AV, Jacobson SG, Drack AV, Ho AC, Charng J, Garafalo AV, et al. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. Nat Med. 2019;25(2):225-8.

98. Slijkerman R, van Diepen H, Albert S, Dona M, Venselaar H, Zang J, et al. Antisense oligonucleotide-based treatment of retinitis pigmentosa caused by mutations in USH2A exon 13. bioRxiv. 2020.

99. Gemayel MC, Bhatwadekar AD, Ciulla T. RNA therapeutics for retinal diseases. Expert Opinion on Biological Therapy. 2020:1-11.

100. Nagel-Wolfrum K, Moeller F, Penner I, Wolfrum U. Translational read-through as an alternative approach for ocular gene therapy of retinal dystrophies caused by in-frame nonsense mutations. Visual neuroscience. 2014;31(4-5):309-16.

101. Ledford H. CRISPR treatment inserted directly into the body for first time. Natur. 2020;579(7798):185-.

102. Fuster-Garcia C, Garcia-Garcia G, Gonzalez-Romero E, Jaijo T, Sequedo MD, Ayuso C, et al. USH2A Gene Editing Using the CRISPR System. Mol Ther Nucleic Acids. 2017;8:529-41.

103. Margulies C, Gloskowski S, Pattali R, Tabbaa D, Giannoukos G, Maeder M, et al., editors. Developing a CRISPR/Cas9 Editing Approach for the Treatment of USH2A-Related Inherited Retinal Degeneration. European Society of Gene and Cell Therapy; 2019; Barcelona, Spain.

104. Zhang Y, Tian Z, Yuan J, Liu C, Liu HL, Ma SQ, et al. The Progress of Gene Therapy for Leber's Optic Hereditary Neuropathy. Curr Gene Ther. 2017;17(4):320-6.

105. Fraldi A, Serafini M, Sorrentino NC, Gentner B, Aiuti A, Bernardo ME. Gene therapy for mucopolysaccharidoses: in vivo and ex vivo approaches. Ital J Pediatr. 2018;44(Suppl 2):130.

106. Zarouchlioti C, Sanchez-Pintado B, Hafford Tear NJ, Klein P, Liskova P, Dulla K, et al. Antisense Therapy for a Common Corneal Dystrophy Ameliorates TCF4 Repeat Expansion-Mediated Toxicity. Am J Hum Genet. 2018;102(4):528-39.

107. Jain A, Zode G, Kasetti RB, Ran FA, Yan W, Sharma TP, et al. CRISPR-Cas9-based treatment of myocilin-associated glaucoma. Proc Natl Acad Sci U S A. 2017;114(42):11199-204.

108. Salzman R, Cook F, Hunt T, Malech HL, Reilly P, Foss-Campbell B, et al. Addressing the Value of Gene Therapy and Enhancing Patient Access to Transformative Treatments. Mol Ther. 2018;26(12):2717-26.

109. Hardjasa A, Ling M, Ma K, Yu H. Investigating the effects of DMSO on PCR fidelity using a restriction digest-based method. J Exp Microbiol Immunol. 2010;14:161-4.

110. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc. 2008;3(9):1452-6.

111. Werle E, Schneider C, Renner M, Volker M, Fiehn W. Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Res. 1994;22(20):4354-5.

112. Fasham J, Arno G, Lin S, Xu M, Carss KJ, Hull S, et al. Delineating the expanding phenotype associated with SCAPER gene mutation. Am J Med Genet A. 2019;179(8):1665-71.

113. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.

114. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

115. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747-54.

116. Laver TW, De Franco E, Johnson MB, Patel K, Ellard S, Weedon MN, et al. SavvyCNV: genome-wide CNV calling from off-target reads. BioRxiv. 2019:617605.

117. Kausar T, Bhatti MA, Ali M, Shaikh RS, Ahmed ZM. OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24. Clin Genet. 2013;84(1):91-3.

118. McKay BS. Pigmentation and vision: Is GPR143 in control? J Neurosci Res. 2018.

119. Kruijt CC, de Wit GC, Bergen AA, Florijn RJ, Schalij-Delfos NE, van Genderen MM. The Phenotypic Spectrum of Albinism. Ophthalmology. 2018;125(12):1953-60.

120. Sjöström A, Kraemer M, Ohlsson J, Villarreal G. Subnormal visual acuity syndromes (SVAS): albinism in Swedish 12-13-year-old children. Documenta ophthalmologica. 2001;103(1):35-46.

121. Sjöström A, Kraemer M, Ohlsson J, Garay-Cerro G, Abrahamsson M, Villarreal G. Subnormal visual acuity (svas) and albinism in mexican 12–13-year-old children. Documenta ophthalmologica. 2004;108(1):9-15.

122. Lewis RA. Ocular Albinism, X-Linked. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews((R)). Seattle (WA)1993.

123. Thomas MG, Maconachie G, Hisaund M, Gottlob I. FRMD7-Related Infantile Nystagmus. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews((R)). Seattle (WA)1993.

124. Poulter JA, Al-Araimi M, Conte I, van Genderen MM, Sheridan E, Carr IM, et al. Recessive mutations in SLC38A8 cause foveal hypoplasia and optic nerve misrouting without albinism. The American Journal of Human Genetics. 2013;93(6):1143-50.

125. Lima Cunha D, Arno G, Corton M, Moosajee M. The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. Genes (Basel). 2019;10(12).

126. Hutton SM, Spritz RA. Comprehensive analysis of oculocutaneous albinism among non-Hispanic caucasians shows that OCA1 is the most prevalent OCA type. J Invest Dermatol. 2008;128(10):2442-50.

127. Rooryck C, Morice-Picard F, Elcioglu NH, Lacombe D, Taieb A, Arveiler B. Molecular diagnosis of oculocutaneous albinism: new mutations in the OCA1-4 genes and practical aspects. Pigment Cell Melanoma Res. 2008;21(5):583-7.

128. Gronskov K, Ek J, Brondum-Nielsen K. Oculocutaneous albinism. Orphanet J Rare Dis. 2007;2:43.

129. Simeonov DR, Wang X, Wang C, Sergeev Y, Dolinska M, Bower M, et al. DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics. Hum Mutat. 2013;34(6):827-35.

130. Grønskov K, Ek J, Brondum-Nielsen K. Oculocutaneous albinism. Orphanet journal of rare diseases. 2007;2(1):1-8.

131. King RA, Pietsch J, Fryer JP, Savage S, Brott MJ, Russell-Eggitt I, et al. Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. Hum Genet. 2003;113(6):502-13.

132. O'Gorman L, Norman CS, Michaels L, Newall T, Crosby AH, Mattocks C, et al. A small gene sequencing panel realises a high diagnostic rate in patients with congenital nystagmus following basic phenotyping. Sci Rep. 2019;9(1):13229.

133. Fukai K, Holmes SA, Lucchese NJ, Siu VM, Weleber RG, Schnur RE, et al. Autosomal recessive ocular albinism associated with a functionally significant tyrosinase gene polymorphism. Nat Genet. 1995;9(1):92-5.

134. Hutton SM, Spritz RA. A comprehensive genetic study of autosomal recessive ocular albinism in Caucasian patients. Invest Ophthalmol Vis Sci. 2008;49(3):868-72.

135. Chiang PW, Spector E, Tsai AC. Oculocutaneous albinism spectrum. Am J Med Genet A. 2009;149A(7):1590-1.

136. Oetting WS, Pietsch J, Brott MJ, Savage S, Fryer JP, Summers CG, et al. The R402Q tyrosinase variant does not cause autosomal recessive ocular albinism. Am J Med Genet A. 2009;149A(3):466-9.

137. Gargiulo A, Testa F, Rossi S, Di Iorio V, Fecarotta S, de Berardinis T, et al. Molecular and clinical characterization of albinism in a large cohort of Italian patients. Invest Ophthalmol Vis Sci. 2011;52(3):1281-9.

138. Preising MN, Forster H, Gonser M, Lorenz B. Screening of TYR, OCA2, GPR143, and MC1R in patients with congenital nystagmus, macular hypoplasia, and fundus hypopigmentation indicating albinism. Mol Vis. 2011;17:939-48.

139. Ghodsinejad Kalahroudi V, Kamalidehghan B, Arasteh Kani A, Aryani O, Tondar M, Ahmadipour F, et al. Two novel tyrosinase (TYR) gene mutations with pathogenic impact on oculocutaneous albinism type 1 (OCA1). PLoS One. 2014;9(9):e106656.

140. Mauri L, Manfredini E, Del Longo A, Veniani E, Scarcello M, Terrana R, et al. Clinical evaluation and molecular screening of a large consecutive series of albino patients. J Hum Genet. 2017;62(2):277-90.

141. Lasseaux E, Plaisant C, Michaud V, Pennamen P, Trimouille A, Gaston L, et al. Molecular characterization of a series of 990 index patients with albinism. Pigment Cell Melanoma Res. 2018;31(4):466-74.

142. Chiang P-W, Drautz JM, Tsai AC-H, Spector E, Clericuzio CL. A new hypothesis of OCA1B. American journal of medical genetics Part A. 2008;146(22):2968-70.

143. Kubal A, Dagnelie G, Goldberg M. Ocular albinism with absent foveal pits but without nystagmus, photophobia, or severely reduced vision. J AAPOS. 2009;13(6):610-2.

144. Thomas MG, Maconachie GD, Sheth V, McLean RJ, Gottlob I. Development and clinical utility of a novel diagnostic nystagmus gene panel using targeted next-generation sequencing. European Journal of Human Genetics. 2017;25(6):725-34.

145. Norman CS, O'Gorman L, Gibson J, Pengelly RJ, Baralle D, Ratnayaka JA, et al. Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (TYR) causing hypomorphic oculocutaneous albinism (OCA1B). Sci Rep. 2017;7(1):4415.

146. Gronskov K, Ek J, Sand A, Scheller R, Bygum A, Brixen K, et al. Birth prevalence and mutation spectrum in danish patients with autosomal recessive albinism. Invest Ophthalmol Vis Sci. 2009;50(3):1058-64.

147. Jagirdar K, Smit DJ, Ainger SA, Lee KJ, Brown DL, Chapman B, et al. Molecular analysis of common polymorphisms within the human Tyrosinase locus and genetic association with pigmentation traits. Pigment Cell Melanoma Res. 2014;27(4):552-64.

148. Campbell P, Ellingford JM, Parry NRA, Fletcher T, Ramsden SC, Gale T, et al. Clinical and genetic variability in children with partial albinism. Sci Rep. 2019;9(1):16576.

149. Gronskov K, Jespersgaard C, Bruun GH, Harris P, Brondum-Nielsen K, Andresen BS, et al. A pathogenic haplotype, common in Europeans, causes autosomal recessive albinism and uncovers missing heritability in OCA1. Sci Rep. 2019;9(1):645.

150. Chaki M, Mukhopadhyay A, Ray K. Determination of variants in the 3'-region of the tyrosinase gene requires locus specific amplification. Hum Mutat. 2005;26(1):53-8.

151. Halaban R, Cheng E, Hebert DN. Coexpression of wild-type tyrosinase enhances maturation of temperature-sensitive tyrosinase mutants. J Invest Dermatol. 2002;119(2):481-8.

152. Team RC. R: A language and environment for statistical computing. 2013.

153. Marti A, Lasseaux E, Ezzedine K, Leaute-Labreze C, Boralevi F, Paya C, et al. Lessons of a day hospital: Comprehensive assessment of patients with albinism in a European setting. Pigment Cell Melanoma Res. 2018;31(2):318-29.

154. Monferme S, Lasseaux E, Duncombe-Poulet C, Hamel C, Defoort-Dhellemmes S, Drumare I, et al. Mild form of oculocutaneous albinism type 1: phenotypic analysis of compound heterozygous patients with the R402Q variant of the TYR gene. Br J Ophthalmol. 2019;103(9):1239-47.

155. Tripathi RK, Giebel LB, Strunk KM, Spritz RA. A polymorphism of the human tyrosinase gene is associated with temperature-sensitive enzymatic activity. Gene Expr. 1991;1(2):103-10.

156. Hudjashov G, Villems R, Kivisild T. Global patterns of diversity and selection in human tyrosinase gene. PLoS One. 2013;8(9):e74307.

157. Mondal M, Sengupta M, Ray K. Functional assessment of tyrosinase variants identified in individuals with albinism is essential for unequivocal determination of genotype-to-phenotype correlation. Br J Dermatol. 2016;175(6):1232-42.

158. King RA, Willaert RK, Schmidt RM, Pietsch J, Savage S, Brott MJ, et al. MC1R mutations modify the classic phenotype of oculocutaneous albinism type 2 (OCA2). Am J Hum Genet. 2003;73(3):638-45.

159. Manga P, Kromberg JG, Box NF, Sturm RA, Jenkins T, Ramsay M. Rufous oculocutaneous albinism in southern African Blacks is caused by mutations in the TYRP1 gene. Am J Hum Genet. 1997;61(5):1095-101.

160. Chiang PW, Fulton AB, Spector E, Hisama FM. Synergistic interaction of the OCA2 and OCA3 genes in a family. Am J Med Genet A. 2008;146A(18):2427-30.

161. Sajid Z, Yousaf S, Waryah YM, Mughal TA, Kausar T, Shahzad M, et al. Genetic Causes of Oculocutaneous Albinism in Pakistani Population. Genes (Basel). 2021;12(4).

162. Wei AH, Yang XM, Lian S, Li W. Genetic analyses of Chinese patients with digenic oculocutaneous albinism. Chin Med J (Engl). 2013;126(2):226-30.

163. Thomas MG, Kumar A, Mohammad S, Proudlock FA, Engle EC, Andrews C, et al. Structural grading of foveal hypoplasia using spectral-domain optical coherence tomography: a predictor of visual acuity? Ophthalmology. 2011;118(8):1653-60.

164. Berson JF, Frank DW, Calvo PA, Bieler BM, Marks MS. A common temperature-sensitive allelic form of human tyrosinase is retained in the endoplasmic reticulum at the nonpermissive temperature. J Biol Chem. 2000;275(16):12281-9.

165. Spritz RA, Ho L, Furumura M, Hearing VJ, Jr. Mutational analysis of copper binding by human tyrosinase. J Invest Dermatol. 1997;109(2):207-12.

166. Tripathi RK, Hearing VJ, Urabe K, Aroca P, Spritz RA. Mutational mapping of the catalytic activities of human tyrosinase. J Biol Chem. 1992;267(33):23707-12.

167. Halaban R, Svedine S, Cheng E, Smicun Y, Aron R, Hebert DN. Endoplasmic reticulum retention is a common defect associated with tyrosinase-negative albinism. Proceedings of the National Academy of Sciences. 2000;97(11):5889-94.

168. Dolinska MB, Kus NJ, Farney SK, Wingfield PT, Brooks BP, Sergeev YV. Oculocutaneous albinism type 1: link between mutations, tyrosinase conformational stability, and enzymatic activity. Pigment cell & melanoma research. 2017;30(1):41-52.

169. Toyofuku K, Wada I, Spritz RA, Hearing VJ. The molecular basis of oculocutaneous albinism type 1 (OCA1): sorting failure and degradation of mutant tyrosinases results in a lack of pigmentation. Biochem J. 2001;355(Pt 2):259-69.

170. Chaki M, Sengupta M, Mondal M, Bhattacharya A, Mallick S, Bhadra R, et al. Molecular and functional studies of tyrosinase variants among Indian oculocutaneous albinism type 1 patients. J Invest Dermatol. 2011;131(1):260-2.

171. Wei AH, Zang DJ, Zhang Z, Yang XM, Li W. Prenatal genotyping of four common oculocutaneous albinism genes in 51 Chinese families. J Genet Genomics. 2015;42(6):279-86.

172. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet. 2008;40(7):835-7.

173. Nan H, Kraft P, Hunter DJ, Han J. Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. Int J Cancer. 2009;125(4):909-17.

174. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 2007;39(12):1443-52.

175. Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, et al. A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet. 2007;81(6):1119-32.

176. Hu HH, Guedj M, Descamps V, Jouary T, Bourillon A, Ezzedine K, et al. Assessment of tyrosinase variants and skin cancer risk in a large cohort of French subjects. J Dermatol Sci. 2011;64(2):127-33.

177. Bouchard B, Fuller BB, Vijayasaradhi S, Houghton AN. Induction of pigmentation in mouse fibroblasts by expression of human tyrosinase cDNA. J Exp Med. 1989;169(6):2029-42.

178. K B, Purohit R. Mutational analysis of TYR gene and its structural consequences in OCA1A. Gene. 2013;513(1):184-95.

179. Kobayashi T, Hearing VJ. Direct interaction of tyrosinase with Tyrp1 to form heterodimeric complexes in vivo. J Cell Sci. 2007;120(Pt 24):4261-8.

180. Orlow SJ, Zhou BK, Chakraborty AK, Drucker M, Pifko-Hirst S, Pawelek JM. High-molecular-weight forms of tyrosinase and the tyrosinase-related proteins: evidence for a melanogenic complex. J Invest Dermatol. 1994;103(2):196-201.

181. Adams DR, Menezes S, Jauregui R, Valivullah ZM, Power B, Abraham M, et al. One-year pilot study on the effects of nitisinone on melanin in patients with OCA-1B. JCI Insight. 2019;4(2).

182. Lee H, Scott J, Griffiths H, Self JE, Lotery A. Oral levodopa rescues retinal morphology and visual function in a murine model of human albinism. Pigment Cell Melanoma Res. 2019;32(5):657-71.

183. Hussain R, Bittles AH. The prevalence and demographic characteristics of consanguineous marriages in Pakistan. J Biosoc Sci. 1998;30(2):261-75.

184. Spritz RA. Molecular genetics of oculocutaneous albinism. Semin Dermatol. 1993;12(3):167-72.

185. Ribero S, Carrera C, Tell-Marti G, Pastorino C, Badenas C, Garcia A, et al. Amelanotic melanoma in oculocutaneous albinism: a genetic, dermoscopic and reflectance confocal microscopy study. Br J Dermatol. 2017;177(6):e333-e5.

186. Shahzad M, Yousaf S, Waryah YM, Gul H, Kausar T, Tariq N, et al. Molecular outcomes, clinical consequences, and genetic diagnosis of Oculocutaneous Albinism in Pakistani population. Scientific reports. 2017;7(1):1-15.

187. Ceyhan-Birsoy O, Murry JB, Machini K, Lebo MS, Yu TW, Fayer S, et al. Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. Am J Hum Genet. 2019;104(1):76-93.

188. Gul H, Shah AH, Harripaul R, Mikhailov A, Prajapati K, Khan E, et al. Genetic studies of multiple consanguineous Pakistani families segregating oculocutaneous albinism identified novel and reported mutations. Ann Hum Genet. 2019;83(4):278-84.

189. Zhong Z, Gu L, Zheng X, Ma N, Wu Z, Duan J, et al. Comprehensive analysis of spectral distribution of a large cohort of Chinese patients with non-syndromic oculocutaneous albinism facilitates genetic diagnosis. Pigment Cell Melanoma Res. 2019;32(5):672-86.

190. Hou YC, Yu HC, Martin R, Cirulli ET, Schenker-Ahmed NM, Hicks M, et al. Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. Proc Natl Acad Sci U S A. 2020;117(6):3053-62.

191. Sajid Z, Yousaf S, Waryah YM, Mughal TA, Kausar T, Shahzad M, et al. Genetic Causes of Oculocutaneous Albinism in Pakistani Population. Genes. 2021;12(4):492.

192. Gul H, Ali MZ, Khan E, Zubair M, Badar M, Khan S, et al. Ophthalmo-genetic analysis of Pakistani patients with nonsyndromic oculocutaneous albinism through whole exome sequencing. J Pak Med Assoc. 2017;67:790-2.

193. Jaworek TJ, Kausar T, Bell SM, Tariq N, Maqsood MI, Sohail A, et al. Molecular genetic studies and delineation of the oculocutaneous albinism phenotype in the Pakistani population. Orphanet J Rare Dis. 2012;7:44.

194. Tripathi RK, Bundey S, Musarella MA, Droetto S, Strunk KM, Holmes SA, et al. Mutations of the tyrosinase gene in Indo-Pakistani patients with type I (tyrosinase-deficient) oculocutaneous albinism (OCA). Am J Hum Genet. 1993;53(6):1173-9.

195. Chong JX, Ouwenga R, Anderson RL, Waggoner DJ, Ober C. A population-based study of autosomal-recessive disease-causing mutations in a founder population. Am J Hum Genet. 2012;91(4):608-20.

196. Oetting WS, Fryer JP, King RA. Mutations of the human tyrosinase gene associated with tyrosinase related oculocutaneous albinism (OCA1). Mutations in brief no. 204. Online. Hum Mutat. 1998;12(6):433-4.

197. Lin Y, Chen X, Yang Y, Che F, Zhang S, Yuan L, et al. Mutational Analysis of TYR, OCA2, and SLC45A2 Genes in Chinese Families with Oculocutaneous Albinism. Mol Genet Genomic Med. 2019;7(7):e00687.

198. Tripathi RK, Strunk KM, Giebel LB, Weleber RG, Spritz RA. Tyrosinase gene mutations in type I (tyrosinase-deficient) oculocutaneous albinism define two clusters of missense substitutions. Am J Med Genet. 1992;43(5):865-71.

199. Nakamura E, Miyamura Y, Matsunaga J, Kano Y, Dakeishi-Hara M, Tanita M, et al. A novel mutation of the tyrosinase gene causing oculocutaneous albinism type 1 (OCA1). J Dermatol Sci. 2002;28(2):102-5.

200. Renugadevi K, Mary J, Perumalsamy V, Seshadri S, Jagadeesh S, Suresh B. Molecular Genetic Testing for Carrier-Prenatal Diagnosis and Computational Analysis of Oculocutaneous Albinism Type 1. J Genet Disor Genet Rep 3. 2014;2:2.

201. Takizawa Y, Kato S, Matsunaga J, Aozaki R, Tomita Y, Nishikawa T, et al. Electron microscopic DOPA reaction test for oculocutaneous albinism. Arch Dermatol Res. 2000;292(6):301-5.

202. Wang X, Zhu Y, Shen N, Peng J, Wang C, Liu H, et al. Mutation analysis of a Chinese family with oculocutaneous albinism. Oncotarget. 2016;7(51):84981-8.

203. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genetics in Medicine. 2018;20(4):435-43.

204. Rego S, Dagan-Rosenfeld O, Zhou W, Sailani MR, Limcaoco P, Colbert E, et al. High-frequency actionable pathogenic exome variants in an average-risk cohort. Molecular Case Studies. 2018;4(6):a003178.

205. Bibi N, Ullah A, Darwesh L, Khan W, Khan T, Ullah K, et al. Identification and Computational Analysis of Novel TYR and SLC45A2 Gene Mutations in Pakistani Families With Identical Non-syndromic Oculocutaneous Albinism. Front Genet. 2020;11:749.

206. King R, Mentink M, Oetting WS. Non-random distribution of missense mutations within the human tyrosinase gene in type I (tyrosinase-related) oculocutaneous albinism. Molecular biology & medicine. 1991;8(1):19-29.

207. Chaki M, Sengupta M, Mondal M, Bhattacharya A, Mallick S. Molecular and functional studies of tyrosinase variants among Indian oculocutaneous albinism type 1 patients. Journal of investigative dermatology. 2011;131(1):260-2.

208. Spritz RA, Lee ST, Fukai K, Brondum-Nielsen K, Chitayat D, Lipson MH, et al. Novel mutations of the P gene in type II oculocutaneous albinism (OCA2). Hum Mutat. 1997;10(2):175-7.

209. Yang L, Liu X, Li Z, Zhang P, Wu B, Wang H, et al. Genetic aetiology of early infant deaths in a neonatal intensive care unit. J Med Genet. 2020;57(3):169-77.

210. Oetting WS, Gardner JM, Fryer JP, Ching A, Durham-Pierre D, King RA, et al. Mutations of the human P gene associated with Type II oculocutaneous albinism (OCA2). Mutations in brief no. 205. Online. Hum Mutat. 1998;12(6):434.

211. Zhang L, Xu B, Zhong Y, Chen X, Zheng H, Jiang W, et al. [A de novo mutation of P gene causes oculocutaneous albinism type 2 with prenatal diagnosis]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2013;30(3):318-21.

212. Lee ST, Nicholls RD, Bundey S, Laxova R, Musarella M, Spritz RA. Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. N Engl J Med. 1994;330(8):529-34.

213. Norman CS, O'Gorman L, Gibson J, Pengelly RJ, Baralle D, Ratnayaka JA, et al. Identification of a functionally significant tri-allelic genotype in the Tyrosinase gene (TYR) causing hypomorphic oculocutaneous albinism (OCA1B). Scientific reports. 2017;7(1):1-9.

214. Campbell P, Ellingford JM, Parry NR, Fletcher T, Ramsden SC, Gale T, et al. Clinical and genetic variability in children with partial albinism. Scientific reports. 2019;9(1):1-10.

215. Mondal M, Sengupta M, Samanta S, Sil A, Ray K. Molecular basis of albinism in India: Evaluation of seven potential candidate genes and some new findings. Gene. 2012;511(2):470-4.

216. UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49(D1):D480-D9.

217. Frints SGM, Hennig F, Colombo R, Jacquemont S, Terhal P, Zimmerman HH, et al. Deleterious de novo variants of X-linked ZC4H2 in females cause a variable phenotype with neurogenic arthrogryposis multiplex congenita. Hum Mutat. 2019;40(12):2270-85.

218. Kamaraj B, Purohit R. Mutational analysis of oculocutaneous albinism: a compact review. Biomed Res Int. 2014;2014:905472.

219. Rehman AU, Santos-Cortez RL, Drummond MC, Shahzad M, Lee K, Morell RJ, et al. Challenges and solutions for gene identification in the presence of familial locus heterogeneity. Eur J Hum Genet. 2015;23(9):1207-15.

220. King RA, Willaert RK, Schmidt RM, Pietsch J, Savage S, Brott MJ, et al. MC1R mutations modify the classic phenotype of oculocutaneous albinism type 2 (OCA2). The American Journal of Human Genetics. 2003;73(3):638-45.

221. Chiang PW, Fulton AB, Spector E, Hisama FM. Synergistic interaction of the OCA2 and OCA3 genes in a family. American Journal of Medical Genetics Part A. 2008;146(18):2427-30.

222. Stevens G, Van Beukering J, Jenkins T, Ramsay M. An intragenic deletion of the P gene is the common mutation causing tyrosinase-positive oculocutaneous albinism in southern African Negroids. American journal of human genetics. 1995;56(3):586.

223. Miyamura Y, Verma IC, Saxena R, Hoshi M, Murase A, Nakamura E, et al. Five novel mutations in tyrosinase gene of Japanese and Indian patients with oculocutaneous albinism type I (OCA1). J Invest Dermatol. 2005;125(2):397-8.

224. Chaki M, Sengupta M, Mukhopadhyay A, Subba Rao I, Majumder PP, Das M, et al. OCA1 in different ethnic groups of india is primarily due to founder mutations in the tyrosinase gene. Annals of human genetics. 2006;70(5):623-30.

225. Sundaresan P, Sil AK, Philp AR, Randolph MA, Natchiar G, Namperumalsamy P. Genetic analysis of oculocutaneous albinism type 1 (OCA1) in Indian families: two novel frameshift mutations in the TYR Gene. Mol Vis. 2004;10:1005-10.

226. Tanita M, Matsunaga J, Miyamura Y, Dakeishi M, Nakamura E, Kono M, et al. Polymorphic sequences of the tyrosinase gene: allele analysis on 16 OCA1 patients in Japan indicate that three polymorphic sequences in the tyrosinase gene promoter could be powerful markers for indirect gene diagnosis. Journal of human genetics. 2002;47(1):1-6.

227. Goto M, Sato-Matsumura KC, Sawamura D, Yokota K, Nakamura H, Shimizu H. Tyrosinase gene analysis in Japanese patients with oculocutaneous albinism. J Dermatol Sci. 2004;35(3):215-20.

228. Li HY, Wu WI, Zheng H, Duan HL, Chen Z, Chen LM. [Prenatal gene diagnosis of oculocutaneous albinism type I]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2006;23(3):280-2.

229. Wang Y, Guo X, Li W, Lian S. Four novel mutations of TYR gene in Chinese OCA1 patients. J Dermatol Sci. 2009;53(1):80-1.

230. Wang Y, Wang Z, Chen M, Fan N, Yang J, Liu L, et al. Mutational Analysis of the TYR and OCA2 Genes in Four Chinese Families with Oculocutaneous Albinism. PLoS One. 2015;10(4):e0125651.

231. Park SH, Chae H, Kim Y, Kim M. Molecular analysis of Korean patients with oculocutaneous albinism. Jpn J Ophthalmol. 2012;56(1):98-103.

232. Gershoni-Baruch R, Rosenmann A, Droetto S, Holmes S, Tripathi RK, Spritz RA. Mutations of the tyrosinase gene in patients with oculocutaneous albinism from various ethnic groups in Israel. American journal of human genetics. 1994;54(4):586.

233. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends Genet. 2018;34(9):666-81.

234. Kee HL, Verhey KJ. Molecular connections between nuclear and ciliary import processes. Cilia. 2013;2(1):1-10.

235. Mockel A, Perdomo Y, Stutzmann F, Letsch J, Marion V, Dollfus H. Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. Prog Retin Eye Res. 2011;30(4):258-74.

236. Cornillie FJ, Lauweryns JM, Corbeel L. Atypical bronchial cilia in children with recurrent respiratory tract infections. A comparative ultrastructural study. Pathol Res Pract. 1984;178(6):595-604.

237. Hildebrandt F, Benzing T, Katsanis N. Ciliopathies. New England Journal of Medicine. 2011;364(16):1533-43.

238. Adams M, Smith UM, Logan CV, Johnson CA. Recent advances in the molecular pathology, cell biology and genetics of ciliopathies. J Med Genet. 2008;45(5):257-67.

239. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. Nat Rev Mol Cell Biol. 2017;18(9):533-47.

240. Wingfield JL, Lechtreck KF, Lorentzen E. Trafficking of ciliary membrane proteins by the intraflagellar transport/BBSome machinery. Essays Biochem. 2018;62(6):753-63.

241. Kobayashi D, Takeda H. Ciliary motility: the components and cytoplasmic preassembly mechanisms of the axonemal dyneins. Differentiation. 2012;83(2):S23-9.

242. Hua K, Ferland RJ. Primary cilia proteins: ciliary and extraciliary sites and functions. Cell Mol Life Sci. 2018;75(9):1521-40.

243. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol. 2017;241(2):294-309.

244. Wallmeier J, Nielsen KG, Kuehni CE, Lucas JS, Leigh MW, Zariwala MA, et al. Motile ciliopathies. Nat Rev Dis Primers. 2020;6(1):77.

245. Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: an emerging class of human genetic disorders. Annu Rev Genomics Hum Genet. 2006;7:125-48.

246. Wheway G, Nazlamova L, Hancock JT. Signaling through the primary cilium. Frontiers in cell and developmental biology. 2018;6:8.

247. Bujakowska KM, Liu Q, Pierce EA. Photoreceptor Cilia and Retinal Ciliopathies. Cold Spring Harb Perspect Biol. 2017;9(10).

248. Lambert SR, Lyons CJ. Taylor and Hoyt's pediatric ophthalmology and strabismus: Elsevier; 2017.

249. Hampshire DJ, Ayub M, Springell K, Roberts E, Jafri H, Rashid Y, et al. MORM syndrome (mental retardation, truncal obesity, retinal dystrophy and micropenis), a new autosomal recessive disorder, links to 9q34. Eur J Hum Genet. 2006;14(5):543-8.

250. Bielas SL, Silhavy JL, Brancati F, Kisseleva MV, Al-Gazali L, Sztriha L, et al. Mutations in INPP5E, encoding inositol polyphosphate-5-phosphatase E, link phosphatidyl inositol signaling to the ciliopathies. Nat Genet. 2009;41(9):1032-6.

251. Jacoby M, Cox JJ, Gayral S, Hampshire DJ, Ayub M, Blockmans M, et al. INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. Nat Genet. 2009;41(9):1027-31.

252. Hardee I, Soldatos A, Davids M, Vilboux T, Toro C, David KL, et al. Defective ciliogenesis in INPP5E-related Joubert syndrome. Am J Med Genet A. 2017;173(12):3231-7.

253. Travaglini L, Brancati F, Silhavy J, Iannicelli M, Nickerson E, Elkhartoufi N, et al. Phenotypic spectrum and prevalence of INPP5E mutations in Joubert syndrome and related disorders. Eur J Hum Genet. 2013;21(10):1074-8.

254. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science. 2001;293(5538):2256-9.

255. Kong AM, Speed CJ, O'Malley CJ, Layton MJ, Meehan T, Loveland KL, et al. Cloning and characterization of a 72-kDa inositol-polyphosphate 5-phosphatase localized to the Golgi network. J Biol Chem. 2000;275(31):24052-64.

256. Xu W, Jin M, Hu R, Wang H, Zhang F, Yuan S, et al. The Joubert syndrome protein Inpp5e controls ciliogenesis by regulating phosphoinositides at the apical membrane. Journal of the American Society of Nephrology. 2017;28(1):118-29.

257. Alfares AA. Applying filtration steps to interpret the results of whole-exome sequencing in a consanguineous population to achieve a high detection rate. Int J Health Sci (Qassim). 2018;12(5):35-43.

258. Fleming LR, Doherty DA, Parisi MA, Glass IA, Bryant J, Fischer R, et al. Prospective Evaluation of Kidney Disease in Joubert Syndrome. Clin J Am Soc Nephrol. 2017;12(12):1962-73.

259. Suzuki T, Miyake N, Tsurusaki Y, Okamoto N, Alkindy A, Inaba A, et al. Molecular genetic analysis of 30 families with Joubert syndrome. Clin Genet. 2016;90(6):526-35.

260. Porto FBO, Jones EM, Branch J, Soens ZT, Maia IM, Sena IFG, et al. Molecular Screening of 43 Brazilian Families Diagnosed with Leber Congenital Amaurosis or Early-Onset Severe Retinal Dystrophy. Genes (Basel). 2017;8(12).

261. Stone EM, Andorf JL, Whitmore SS, DeLuca AP, Giacalone JC, Streb LM, et al. Clinically Focused Molecular Investigation of 1000 Consecutive Families with Inherited Retinal Disease. Ophthalmology. 2017;124(9):1314-31.

262. Tsurusaki Y, Kobayashi Y, Hisano M, Ito S, Doi H, Nakashima M, et al. The diagnostic utility of exome sequencing in Joubert syndrome and related disorders. J Hum Genet. 2013;58(2):113-5.

263. Sharon D, Ben-Yosef T, Goldenberg-Cohen N, Pras E, Gradstein L, Soudry S, et al. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). Human mutation. 2020;41(1):140-9.

264. Sönmez F, Güzünler-Şen M, Yılmaz D, Cömertpay G, Heise M, Çırak S, et al. Development of end-stage renal disease at a young age in two cases with Joubert syndrome. Turk J Pediatr. 2014;56:458-61.

265. Shetty M, Ramdas N, Sahni S, Mullapudi N, Hegde S. A Homozygous Missense Variant in INPP5E Associated with Joubert Syndrome and Related Disorders. Mol Syndromol. 2017;8(6):313-7.

266. Wang H, Wang X, Zou X, Xu S, Li H, Soens ZT, et al. Comprehensive Molecular Diagnosis of a Large Chinese Leber Congenital Amaurosis Cohort. Invest Ophthalmol Vis Sci. 2015;56(6):3642-55.

267. Lane KT, Beese LS. Thematic review series: lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type I. Journal of lipid research. 2006;47(4):681-99.

268. Layton MJ, Meehan T, Loveland KL, Cheema S, Kong AM, Speed CJ, et al. Cloning and characterization of a 72-kDa inositol-polyphosphate 5-phosphatase localized to the Golgi network. Journal of Biological Chemistry. 2000;275(31):24052-64.

269. de Goede C, Yue WW, Yan G, Ariyaratnam S, Chandler KE, Downes L, et al. Role of reverse phenotyping in interpretation of next generation sequencing data and a review of INPP5E related disorders. Eur J Paediatr Neurol. 2016;20(2):286-95.

270. Birtel J, Eisenberger T, Gliem M, Muller PL, Herrmann P, Betz C, et al. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. Sci Rep. 2018;8(1):4824.

271. Xu Y, Guan L, Xiao X, Zhang J, Li S, Jiang H, et al. Mutation analysis in 129 genes associated with other forms of retinal dystrophy in 157 families with retinitis pigmentosa based on exome sequencing. Mol Vis. 2015;21:477-86.

272. Sergouniotis PI, Barton SJ, Waller S, Perveen R, Ellingford JM, Campbell C, et al. The role of small in-frame insertions/deletions in inherited eye disorders and how structural modelling can help estimate their pathogenicity. Orphanet journal of rare diseases. 2016;11(1):1-8.

273. Wang X, Wang H, Sun V, Tuan HF, Keser V, Wang K, et al. Comprehensive molecular diagnosis of 179 Leber congenital amaurosis and juvenile retinitis pigmentosa patients by targeted next generation sequencing. J Med Genet. 2013;50(10):674-88.

274. Ben-Salem S, Al-Shamsi AM, Gleeson JG, Ali BR, Al-Gazali L. Mutation spectrum of Joubert syndrome and related disorders among Arabs. Human genome variation. 2014;1(1):1-10.

275. Toma C, Ruberto G, Marzi F, Vandelli G, Signorini S, Valente EM, et al. Macular staphyloma in patients affected by Joubert syndrome with retinal dystrophy: a new finding detected by SD-OCT. Doc Ophthalmol. 2018;137(1):25-36.

276. Bachmann-Gagescu R, Dempsey JC, Phelps IG, O'Roak BJ, Knutzen DM, Rue TC, et al. Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. J Med Genet. 2015;52(8):514-22.

277. Chen F, Sun SZ, Tang HX, Li RP, Wang W, Liu K, et al. [Joubert syndrome caused by INPP5E mutations: report of a family]. Zhongguo Dang Dai Er Ke Za Zhi. 2018;20(10):861-3.

278. Poretti A, Dietrich Alber F, Brancati F, Dallapiccola B, Valente EM, Boltshauser E. Normal cognitive functions in joubert syndrome. Neuropediatrics. 2009;40(6):287-90.

279. Summers AC, Snow J, Wiggs E, Liu AG, Toro C, Poretti A, et al. Neuropsychological phenotypes of 76 individuals with Joubert syndrome evaluated at a single center. American Journal of Medical Genetics Part A. 2017;173(7):1796-812.

280. Klein D, Ammann F. The syndrome of Laurence-Moon-Bardet-Biedl and allied diseases in Switzerland. Clinical, genetic and epidemiological studies. J Neurol Sci. 1969;9(3):479-513.

281. Beales PL, Warner AM, Hitman GA, Thakker R, Flinter FA. Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. J Med Genet. 1997;34(2):92-8.

282. Farag TI, Teebi AS. High incidence of Bardet Biedl syndrome among the Bedouin. Clinical genetics. 1989;36(6):463-4.

283. Moore SJ, Green JS, Fan Y, Bhogal AK, Dicks E, Fernandez BA, et al. Clinical and genetic epidemiology of Bardet-Biedl syndrome in Newfoundland: a 22-year prospective, population-based, cohort study. Am J Med Genet A. 2005;132A(4):352-60.

284. Beales P, Elcioglu N, Woolf A, Parker D, Flinter F. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. Journal of medical genetics. 1999;36(6):437-46.

285. Denniston AK, Beales PL, Tomlins PJ, Good P, Langford M, Foggensteiner L, et al. Evaluation of visual function and needs in adult patients with bardet-biedl syndrome. Retina. 2014;34(11):2282-9.

286. Forsythe E, Kenny J, Bacchelli C, Beales PL. Managing Bardet-Biedl Syndrome-Now and in the Future. Front Pediatr. 2018;6:23.

287. Nachury MV, Loktev AV, Zhang Q, Westlake CJ, Peranen J, Merdes A, et al. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell. 2007;129(6):1201-13.

288. Seo S, Baye LM, Schulz NP, Beck JS, Zhang Q, Slusarski DC, et al. BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins

and mediate BBSome assembly. Proc Natl Acad Sci U S A. 2010;107(4):1488-93.

289. Jin H, White SR, Shida T, Schulz S, Aguiar M, Gygi SP, et al. The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. Cell. 2010;141(7):1208-19.

290. Seo S, Zhang Q, Bugge K, Breslow DK, Searby CC, Nachury MV, et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet. 2011;7(11):e1002358.

291. Stowe TR, Wilkinson CJ, Iqbal A, Stearns T. The centriolar satellite proteins Cep72 and Cep290 interact and are required for recruitment of BBS proteins to the cilium. Mol Biol Cell. 2012;23(17):3322-35.

292. Heon E, Kim G, Qin S, Garrison JE, Tavares E, Vincent A, et al. Mutations in C8ORF37 cause Bardet Biedl syndrome (BBS21). Hum Mol Genet. 2016;25(11):2283-94.

293. Priya S, Nampoothiri S, Sen P, Sripriya S. Bardet–Biedl syndrome: genetics, molecular pathophysiology, and disease management. Indian journal of ophthalmology. 2016;64(9):620.

294. Strauss KA, Puffenberger EG. Genetics, medicine, and the Plain people. Annu Rev Genomics Hum Genet. 2009;10:513-36.

295. Stone DL, Slavotinek A, Bouffard GG, Banerjee-Basu S, Baxevanis AD, Barr M, et al. Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. Nat Genet. 2000;25(1):79-82.

296. Mykytyn K, Nishimura DY, Searby CC, Shastri M, Yen H-j, Beck JS, et al. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. Nature genetics. 2002;31(4):435-8.

297. Mykytyn K, Nishimura DY, Searby CC, Beck G, Bugge K, Haines HL, et al. Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). The American Journal of Human Genetics. 2003;72(2):429-37.

298. Chou H-T, Apelt L, Farrell DP, White SR, Woodsmith J, Svetlov V, et al. The molecular architecture of native BBSome obtained by an integrated structural approach. Structure. 2019;27(9):1384-94. e4.

299. Riise R, Andreasson S, Borgastrom MK, Wright AF, Tommerup N, Rosenberg T, et al. Intrafamilial variation of the phenotype in Bardet-Biedl syndrome. Br J Ophthalmol. 1997;81(5):378-85.

300. Forsythe E, Beales PL. Bardet-Biedl syndrome. Eur J Hum Genet. 2013;21(1):8-13.

301. Niederlova V, Modrak M, Tsyklauri O, Huranova M, Stepanek O. Meta-analysis of genotype-phenotype associations in Bardet-Biedl syndrome uncovers differences among causative genes. Human mutation. 2019;40(11):2068-87.

302. Katsanis N, Eichers ER, Ansley SJ, Lewis RA, Kayserili H, Hoskins BE, et al. BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. The American Journal of Human Genetics. 2002;71(1):22-9.

303. Beales PL, Badano JL, Ross AJ, Ansley SJ, Hoskins BE, Kirsten B, et al. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. The American Journal of Human Genetics. 2003;72(5):1187-99.

304. Badano JL, Kim JC, Hoskins BE, Lewis RA, Ansley SJ, Cutler DJ, et al. Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet–Biedl patients with two mutations at a second BBS locus. Human molecular genetics. 2003;12(14):1651-9.

305. Katsanis N. The oligogenic properties of Bardet-Biedl syndrome. Hum Mol Genet. 2004;13 Spec No 1:R65-71.

306. Castro-Sanchez S, Alvarez-Satta M, Corton M, Guillen E, Ayuso C, Valverde D. Exploring genotype-phenotype relationships in Bardet-Biedl syndrome families. J Med Genet. 2015;52(8):503-13.

307. Stoetzel C, Laurier V, Davis EE, Muller J, Rix S, Badano JL, et al. BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. Nature genetics. 2006;38(5):521-4.

308. Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, et al. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nat Genet. 2008;40(4):443-8.

309. Fan Y, Esmail MA, Ansley SJ, Blacque OE, Boroevich K, Ross AJ, et al. Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. Nat Genet. 2004;36(9):989-93.

310. Hjortshoj TD, Gronskov K, Philp AR, Nishimura DY, Riise R, Sheffield VC, et al. Bardet-Biedl syndrome in Denmark--report of 13 novel sequence variations in six genes. Hum Mutat. 2010;31(4):429-36.

311. Bin J, Madhavan J, Ferrini W, Mok CA, Billingsley G, Heon E. BBS7 and TTC8 (BBS8) mutations play a minor role in the mutational load of Bardet-Biedl syndrome in a multiethnic population. Hum Mutat. 2009;30(7):E737-46.

312. Badano JL, Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, et al. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. Nature. 2006;439(7074):326-30.

313. Gonzalez-Del Pozo M, Mendez-Vidal C, Santoyo-Lopez J, Vela-Boza A, Bravo-Gil N, Rueda A, et al. Deciphering intrafamilial phenotypic variability by exome sequencing in a Bardet-Biedl family. Mol Genet Genomic Med. 2014;2(2):124-33.

314. Khanna H, Davis EE, Murga-Zamalloa CA, Estrada-Cuzcano A, Lopez I, Den Hollander AI, et al. A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. Nature genetics. 2009;41(6):739-45.

315. Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nature genetics. 2011;43(3):189-96.

316. Lindstrand A, Davis EE, Carvalho CM, Pehlivan D, Willer JR, Tsai I-C, et al. Recurrent CNVs and SNVs at the NPHP1 locus contribute pathogenic alleles to Bardet-Biedl syndrome. The American Journal of Human Genetics. 2014;94(5):745-54.

317. Zhang Y, Seo S, Bhattarai S, Bugge K, Searby CC, Zhang Q, et al. BBS mutations modify phenotypic expression of CEP290-related ciliopathies. Hum Mol Genet. 2014;23(1):40-51.

318. Zaghloul NA, Katsanis N. Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. The Journal of clinical investigation. 2009;119(3):428-37.

319. Estrada-Cuzcano A, Koenekoop RK, Senechal A, De Baere EB, de Ravel T, Banfi S, et al. BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. Arch Ophthalmol. 2012;130(11):1425-32.

320. Yildiz Bolukbasi E, Mumtaz S, Afzal M, Woehlbier U, Malik S, Tolun A. Homozygous mutation in CEP19, a gene mutated in morbid obesity, in Bardet-Biedl syndrome with predominant postaxial polydactyly. J Med Genet. 2018;55(3):189-97.

321. Davis RE, Swiderski RE, Rahmouni K, Nishimura DY, Mullins RF, Agassandian K, et al. A knockin mouse model of the Bardet–Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. Proceedings of the National Academy of Sciences. 2007;104(49):19422-7.

322. Zaghloul NA, Liu Y, Gerdes JM, Gascue C, Oh EC, Leitch CC, et al. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Proc Natl Acad Sci U S A. 2010;107(23):10602-7.

323. Azari AA, Aleman TS, Cideciyan AV, Schwartz SB, Windsor EA, Sumaroka A, et al. Retinal disease expression in Bardet-Biedl syndrome-1 (BBS1) is a spectrum from maculopathy to retina-wide degeneration. Investigative ophthalmology & visual science. 2006;47(11):5004-10.

324. Cannon PS, Clayton-Smith J, Beales PL, Lloyd IC. Bardet-biedl syndrome: an atypical phenotype in brothers with a proven BBS1 mutation. Ophthalmic Genet. 2008;29(3):128-32.

325. O'Sullivan J, Mullaney BG, Bhaskar SS, Dickerson JE, Hall G, O'Grady A, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet. 2012;49(5):322-6.

326. Wang X, Wang H, Sun V, Tuan H-F, Keser V, Wang K, et al. Comprehensive molecular diagnosis of 179 Leber congenital amaurosis and juvenile retinitis pigmentosa patients by targeted next generation sequencing. Journal of medical genetics. 2013;50(10):674-88.

327. Hichri H, Stoetzel C, Laurier V, Caron S, Sigaudy S, Sarda P, et al. Testing for triallelism: analysis of six BBS genes in a Bardet-Biedl syndrome family cohort. Eur J Hum Genet. 2005;13(5):607-16.

328. Fauser S, Munz M, Besch D. Further support for digenic inheritance in Bardet-Biedl syndrome. Journal of medical genetics. 2003;40(8):e104-e.

329. Smaoui N, Chaabouni M, Sergeev YV, Kallel H, Li S, Mahfoudh N, et al. Screening of the eight BBS genes in Tunisian families: no evidence of triallelism. Investigative ophthalmology & visual science. 2006;47(8):3487-95.

330. Abu-Safieh L, Al-Anazi S, Al-Abdi L, Hashem M, Alkuraya H, Alamr M, et al. In search of triallelism in Bardet-Biedl syndrome. Eur J Hum Genet. 2012;20(4):420-7.

331. Laurier V, Stoetzel C, Muller J, Thibault C, Corbani S, Jalkh N, et al. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. Eur J Hum Genet. 2006;14(11):1195-203.

332. Croft JB, Morrell D, Chase CL, Swift M. Obesity in heterozygous carriers of the gene for the Bardet-Biedl syndrome. Am J Med Genet. 1995;55(1):12-5.

333. Kim LS, Fishman GA, Seiple WH, Szlyk JP, Stone EM. Retinal dysfunction in carriers of bardet-biedl syndrome. Ophthalmic Genet. 2007;28(3):163-8.

334. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet. 1999;36(6):437-46.

335. Croft JB, Swift M. Obesity, hypertension, and renal disease in relatives of Bardet-Biedl syndrome sibs. Am J Med Genet. 1990;36(1):37-42.

336. Webb MP, Dicks EL, Green JS, Moore SJ, Warden GM, Gamberg JS, et al. Autosomal recessive Bardet–Biedl syndrome: first-degree relatives have no predisposition to metabolic and renal disorders. Kidney international. 2009;76(2):215-23.

337. Shamseldin HE, Shaheen R, Ewida N, Bubshait DK, Alkuraya H, Almardawi E, et al. The morbid genome of ciliopathies: an update. Genet Med. 2020;22(6):1051-60.

338. Lindstrand A, Frangakis S, Carvalho CM, Richardson EB, McFadden KA, Willer JR, et al. Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. Am J Hum Genet. 2016;99(2):318-36.

339. Tsang WY, Wang L, Chen Z, Sánchez I, Dynlacht BD. SCAPER, a novel cyclin A–interacting protein that regulates cell cycle progression. The Journal of cell biology. 2007;178(4):621-33.

340. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature. 2011;478(7367):57-63.

341. Tatour Y, Sanchez-Navarro I, Chervinsky E, Hakonarson H, Gawi H, Tahsin-Swafiri S, et al. Mutations in SCAPER cause autosomal recessive retinitis pigmentosa with intellectual disability. Journal of medical genetics. 2017.

342. Wormser O, Gradstein L, Yogev Y, Perez Y, Kadir R, Goliand I, et al. SCAPER localizes to primary cilia and its mutation affects cilia length, causing Bardet-Biedl syndrome. Eur J Hum Genet. 2019.

343. Jauregui R, Thomas AL, Liechty B, Velez G, Mahajan VB, Clark L, et al. SCAPER-associated nonsyndromic autosomal recessive retinitis pigmentosa. American journal of medical genetics Part A. 2019;179(2):312-6.

344. Hu H, Kahrizi K, Musante L, Fattahi Z, Herwig R, Hosseini M, et al. Genetics of intellectual disability in consanguineous families. Mol Psychiatry. 2018.

345. Kahrizi K, Huber M, Galetzka D, Dewi S, Schroder J, Weis E, et al. Homozygous variants in the gene SCAPER cause syndromic intellectual disability. Am J Med Genet A. 2019;179(7):1214-25.

346. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Human mutation. 2015;36(10):928-30.

347. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Human mutation. 2015;36(10):915-21.

348. Turro E, Astle WJ, Megy K, Graf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. Nature. 2020;583(7814):96-102.

349. Cole TJ, Williams AF, Wright CM. Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. Annals of human biology. 2011;38(1):7-11.

350. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Statistics in medicine. 1998;17(4):407-29.

351. Pruett RC. Retinitis pigmentosa: clinical observations and correlations. Trans Am Ophthalmol Soc. 1983;81:693-735.

352. Goldmann T, Overlack N, Wolfrum U, Nagel-Wolfrum K. PTC124-mediated translational readthrough of a nonsense mutation causing Usher syndrome type 1C. Human gene therapy. 2011;22(5):537-47.

353. Schwarz N, Carr A-J, Lane A, Moeller F, Chen LL, Aguilà M, et al. Translational read-through of the RP2 Arg120stop mutation in patient iPSC-derived retinal pigment epithelium cells. Human molecular genetics. 2015;24(4):972-86.

354. Castro-Sánchez S, Suarez-Bregua P, Novas R, Álvarez-Satta M, Badano JL, Rotllant J, et al. Functional analysis of new human Bardet-Biedl syndrome loci specific variants in the zebrafish model. Scientific reports. 2019;9(1):1-12.

355. Norris DP, Grimes DT. Mouse models of ciliopathies: the state of the art. Disease models & mechanisms. 2012;5(3):299-312.

356. Baker K, Beales PL, editors. Making sense of cilia in disease: the human ciliopathies. American journal of medical genetics part C: seminars in medical genetics; 2009: Wiley Online Library.

357. Seo S, Mullins RF, Dumitrescu AV, Bhattarai S, Gratie D, Wang K, et al. Subretinal gene therapy of mice with Bardet-Biedl syndrome type 1. Investigative ophthalmology & visual science. 2013;54(9):6118-32.

358. Simons DL, Boye SL, Hauswirth WW, Wu SM. Gene therapy prevents photoreceptor death and preserves retinal function in a Bardet-Biedl syndrome mouse model. Proceedings of the National Academy of Sciences. 2011;108(15):6276-81.

359. Drack AV, Dumitrescu AV, Bhattarai S, Gratie D, Stone EM, Mullins R, et al. TUDCA slows retinal degeneration in two different mouse models of retinitis pigmentosa and prevents obesity in Bardet-Biedl syndrome type 1 mice. Investigative ophthalmology & visual science. 2012;53(1):100-6.

360. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health. 2015;18(6):906-14.

361. Sobrido MJ, Bauer P, de Koning T, Klopstock T, Nadjar Y, Patterson MC, et al. Recommendations for patient screening in ultra-rare inherited metabolic diseases: what have we learned from Niemann-Pick disease type C? Orphanet J Rare Dis. 2019;14(1):20.

362. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28(2):165-73.

363. Devuyst O, Knoers NV, Remuzzi G, Schaefer F, Board of the Working Group for Inherited Kidney Diseases of the European Renal A, European D, et al. Rare inherited kidney diseases: challenges, opportunities, and perspectives. Lancet. 2014;383(9931):1844-59.

364. Hennekam RC. Care for patients with ultra-rare disorders. Eur J Med Genet. 2011;54(3):220-4.

365. Nikopoulos K, Cisarova K, Quinodoz M, Koskiniemi-Kuendig H, Miyake N, Farinelli P, et al. A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy. Nature communications. 2019;10(1):1-7.

366. Miraldi Utz V, Coussa RG, Antaki F, Traboulsi EI. Gene therapy for RPE65-related retinal disease. Ophthalmic Genet. 2018;39(6):671-7.

367. Gorman GS, Schaefer AM, Ng Y, Gomez N, Blakely EL, Alston CL, et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. Ann Neurol. 2015;77(5):753-9.

368. Alston CL, Davison JE, Meloni F, van der Westhuizen FH, He L, Hornig-Do HT, et al. Recessive germline SDHA and SDHB mutations causing leukodystrophy and isolated mitochondrial complex II deficiency. J Med Genet. 2012;49(9):569-77.

369. Hoekstra AS, Bayley JP. The role of complex II in disease. Biochim Biophys Acta. 2013;1827(5):543-51.

370. van den Heuvel L, Smeitink J. The oxidative phosphorylation (OXPHOS) system: nuclear genes and human genetic diseases. Bioessays. 2001;23(6):518-25.

371. Renkema GH, Wortmann SB, Smeets RJ, Venselaar H, Antoine M, Visser G, et al. SDHA mutations causing a multisystem mitochondrial disease: novel mutations and genetic overlap with hereditary tumors. Eur J Hum Genet. 2015;23(2):202-9.

372. Jackson CB, Nuoffer JM, Hahn D, Prokisch H, Haberberger B, Gautschi M, et al. Mutations in SDHD lead to autosomal recessive encephalomyopathy and isolated mitochondrial complex II deficiency. J Med Genet. 2014;51(3):170-5.

373. Alston CL, Ceccatelli Berti C, Blakely EL, Olahova M, He L, McMahon CJ, et al. A recessive homozygous p.Asp92Gly SDHD mutation causes prenatal cardiomyopathy and a severe mitochondrial complex II deficiency. Hum Genet. 2015;134(8):869-79.

374. Ghezzi D, Goffrini P, Uziel G, Horvath R, Klopstock T, Lochmuller H, et al. SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy. Nat Genet. 2009;41(6):654-6.

375. Jain-Ghai S, Cameron JM, Al Maawali A, Blaser S, MacKay N, Robinson B, et al. Complex II deficiency--a case report and review of the literature. Am J Med Genet A. 2013;161A(2):285-94.

376. Chinnery PF, Schon EA. Mitochondria. J Neurol Neurosurg Psychiatry. 2003;74(9):1188-99.

377. Ostergaard E, Bradinova I, Ravn SH, Hansen FJ, Simeonov E, Christensen E, et al. Hypertrichosis in patients with SURF1 mutations. Am J Med Genet A. 2005;138(4):384-8.

378. Falk MJ. Neurodevelopmental manifestations of mitochondrial disease. J Dev Behav Pediatr. 2010;31(7):610-21.

379. Baysal BE, Ferrell RE, Willett-Brozick JE, Lawrence EC, Myssiorek D, Bosch A, et al. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science. 2000;287(5454):848-51.

380. Hensen EF, Siemers MD, Jansen JC, Corssmit EP, Romijn JA, Tops CM, et al. Mutations in SDHD are the major determinants of the clinical characteristics of Dutch head and neck paraganglioma patients. Clin Endocrinol (Oxf). 2011;75(5):650-5.

381. van der Mey AG, Maaswinkel-Mooy PD, Cornelisse CJ, Schmidt PH, van de Kamp JJ. Genomic imprinting in hereditary glomus tumours: evidence for new genetic theory. Lancet. 1989;2(8675):1291-4.

382. Yeap PM, Tobias ES, Mavraki E, Fletcher A, Bradshaw N, Freel EM, et al. Molecular analysis of pheochromocytoma after maternal transmission of SDHD mutation elucidates mechanism of parent-of-origin effect. J Clin Endocrinol Metab. 2011;96(12):E2009-13.

383. Hensen EF, Jordanova ES, van Minderhout IJ, Hogendoorn PC, Taschner PE, van der Mey AG, et al. Somatic loss of maternal chromosome 11 causes parent-of-origin-dependent inheritance in SDHD-linked paraganglioma and phaeochromocytoma families. Oncogene. 2004;23(23):4076-83.

384. Hensen EF, van Duinen N, Jansen JC, Corssmit EP, Tops CM, Romijn JA, et al. High prevalence of founder mutations of the succinate dehydrogenase genes in the Netherlands. Clin Genet. 2012;81(3):284-8.

385. Grønborg S, Darin N, Miranda MJ, Damgaard B, Cayuela JA, Oldfors A, et al. Leukoencephalopathy due to complex II deficiency and Bi-Allelic SDHB mutations: further cases and implications for genetic counselling. JIMD Reports, Volume 33: Springer; 2016. p. 69-77.

386. Khandekar R, Kishore H, Mansu RM, Awan H. The status of childhood blindness and functional low vision in the Eastern Mediterranean region in 2012. Middle East Afr J Ophthalmol. 2014;21(4):336-43.

387. Kazmi HS, Shah AA, Awan AA, Khan J, Siddiqui N. Status of children in blind schools in the northern areas of Pakistan. J Ayub Med Coll Abbottabad. 2007;19(4):37-9.

388. Chen J, Ma Z, Jiao X, Fariss R, Kantorow WL, Kantorow M, et al. Mutations in FYCO1 cause autosomal-recessive congenital cataracts. Am J Hum Genet. 2011;88(6):827-38.

389. Chen J, Wang Q, Cabrera PE, Zhong Z, Sun W, Jiao X, et al. Molecular Genetic Analysis of Pakistani Families With Autosomal Recessive Congenital Cataracts by Homozygosity Screening. Invest Ophthalmol Vis Sci. 2017;58(4):2207-17.

390. Stoilov I, Akarsu AN, Alozie I, Child A, Barsoum-Homsy M, Turacli ME, et al. Sequence analysis and homology modeling suggest that primary congenital glaucoma on 2p21 results from mutations disrupting either the hinge region or the conserved core structures of cytochrome P4501B1. The American Journal of human genetics. 1998;62(3):573-84.

391. Rauf B, Irum B, Kabir F, Firasat S, Naeem MA, Khan SN, et al. A spectrum of CYP1B1 mutations associated with primary congenital glaucoma in families of Pakistani descent. Hum Genome Var. 2016;3:16021.

392. Sheikh SA, Waryah AM, Narsani AK, Shaikh H, Gilal IA, Shah K, et al. Mutational spectrum of the CYP1B1 gene in Pakistani patients with primary congenital glaucoma: novel variants and genotype-phenotype correlations. Mol Vis. 2014;20:991-1001.

393. Kondo-Saitoh A, Matsumoto N, Sasaki T, Egashira M, Saitoh A, Yamada K, et al. Two nonsense mutations of PAX6 in two Japanese aniridia families: case report and review of the literature. Eur J Ophthalmol. 2000;10(2):167-72.

394. Abouzeid H, Youssef MA, ElShakankiri N, Hauser P, Munier FL, Schorderet DF. PAX6 aniridia and interhemispheric brain anomalies. Mol Vis. 2009;15:2074-83.

395. Chograni M, Derouiche K, Chaabouni M, Lariani I, Bouhamed HC. Molecular analysis of the PAX6 gene for aniridia and congenital cataracts in Tunisian families. Hum Genome Var. 2014;1:14008.

396. Primignani P, Allegrini D, Manfredini E, Romitti L, Mauri L, Patrosso MC, et al. Screening of PAX6 gene in Italian congenital aniridia patients revealed four novel mutations. Ophthalmic Genet. 2016;37(3):307-13.

397. Lin Y, Gao H, Zhu Y, Chen C, Li T, Liu B, et al. Two Paired Box 6 mutations identified in Chinese patients with classic congenital aniridia and cataract. Mol Med Rep. 2018;18(5):4439-45.

398. Syrimis A, Nicolaou N, Alexandrou A, Papaevripidou I, Nicolaou M, Loukianou E, et al. Molecular analysis of Cypriot families with aniridia reveals a novel PAX6 mutation. Mol Med Rep. 2018;18(2):1623-7.

399. Cross E, Duncan-Flavell PJ, Howarth RJ, Crooks RO, Thomas NS, Bunyan DJ. Screening of a large PAX6 cohort identified many novel variants and emphasises the importance of the paired and homeobox domains. Eur J Med Genet. 2020;63(7):103940.

400. Tian W, Zhu XR, Qiao CY, Ma YN, Yang FY, Zhou Z, et al. Heterozygous PAX6 mutations may lead to hyper-proinsulinaemia and glucose intolerance: A case-control study in families with congenital aniridia. Diabet Med. 2021;38(2):e14456.

401. Oh J, Ho L, Ala-Mello S, Amato D, Armstrong L, Bellucci S, et al. Mutation analysis of patients with Hermansky-Pudlak syndrome: a frameshift hot spot in the HPS gene and apparent locus heterogeneity. Am J Hum Genet. 1998;62(3):593-8.

402. Wei A, Yuan Y, Bai D, Ma J, Hao Z, Zhang Y, et al. NGS-based 100-gene panel of hypopigmentation identifies mutations in Chinese Hermansky-Pudlak syndrome patients. Pigment Cell Melanoma Res. 2016;29(6):702-6.

403. Theunissen TEJ, Sallevelt S, Hellebrekers D, de Koning B, Hendrickx ATM, van den Bosch BJC, et al. Rapid Resolution of Blended or Composite Multigenic Disease in Infants by Whole-Exome Sequencing. J Pediatr. 2017;182:371-4 e2.

404. Okamura K, Hayashi M, Abe Y, Kono M, Nakajima K, Aoyama Y, et al. NGS-based targeted resequencing identified rare subtypes of albinism: Providing accurate molecular diagnosis for Japanese patients with albinism. Pigment Cell Melanoma Res. 2019;32(6):848-53.

405. Huizing M, Malicdan MCV, Wang JA, Pri-Chen H, Hess RA, Fischer R, et al. Hermansky-Pudlak syndrome: Mutation update. Hum Mutat. 2020;41(3):543-80.

406. Sim W, Kim SY, Han J, Rim TH, Lee JG, Paik HC, et al. Extracorporeal membrane oxygenation bridge to lung transplantation in a patient with Hermansky-Pudlak syndrome and progressive pulmonary fibrosis. Acute and critical care. 2019;34(1):95.

407. Ito S, Suzuki T, Inagaki K, Suzuki N, Takamori K, Yamada T, et al. High frequency of Hermansky-Pudlak syndrome type 1 (HPS1) among Japanese albinism patients and functional analysis of HPS1 mutant protein. J Invest Dermatol. 2005;125(4):715-20.

408. Kanazu M, Arai T, Sugimoto C, Kitaichi M, Akira M, Abe Y, et al. An intractable case of Hermansky-Pudlak syndrome. Intern Med. 2014;53(22):2629-34.

409. Moretti A, Li J, Donini S, Sobol RW, Rizzi M, Garavaglia S. Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD(+) and retinoic acid. Sci Rep. 2016;6:35710.

410. Verma AS, Fitzpatrick DR. Anophthalmia and microphthalmia. Orphanet J Rare Dis. 2007;2:47.

411. Bardakjian TM, Schneider A. The genetics of anophthalmia and microphthalmia. Curr Opin Ophthalmol. 2011;22(5):309-13.

412. Fares-Taie L, Gerber S, Chassaing N, Clayton-Smith J, Hanein S, Silva E, et al. ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. Am J Hum Genet. 2013;92(2):265-70.

413. Ragge NK, Subak-Sharpe ID, Collin JR. A practical guide to the management of anophthalmia and microphthalmia. Eye (Lond). 2007;21(10):1290-300.

414. Jimenez NL, Flannick J, Yahyavi M, Li J, Bardakjian T, Tonkin L, et al. Targeted 'next-generation' sequencing in anophthalmia and microphthalmia patients confirms SOX2, OTX2 and FOXE3 mutations. BMC Med Genet. 2011;12:172.

415. Mory A, Ruiz FX, Dagan E, Yakovtseva EA, Kurolap A, Pares X, et al. A missense mutation in ALDH1A3 causes isolated microphthalmia/anophthalmia in nine individuals from an inbred Muslim kindred. Eur J Hum Genet. 2014;22(3):419-22.

416. Aldahmesh MA, Khan AO, Hijazi H, Alkuraya FS. Mutations in ALDH1A3 cause microphthalmia. Clin Genet. 2013;84(2):128-31.

417. Roos L, Fang M, Dali C, Jensen H, Christoffersen N, Wu B, et al. A homozygous mutation in a consanguineous family consolidates the role of ALDH1A3 in autosomal recessive microphthalmia. Clin Genet. 2014;86(3):276-81.

418. Dehghani M, Dehghan Tezerjani M, Metanat Z, Vahidi Mehrjardi MY. A Novel Missense Mutation in the ALDH13 Gene Causes Anophthalmia in Two Unrelated Iranian Consanguineous Families. Int J Mol Cell Med. 2017;6(2):131-4.

419. Alabdullatif M, Al Dhaibani MA, Khassawneh MY, El-Hattab AW. Chromosomal microarray in a highly consanguineous population: diagnostic yield, utility of regions of homozygosity, and novel mutations. Clinical genetics. 2017;91(4):616-22.

420. Ullah E, Nadeem Saqib MA, Sajid S, Shah N, Zubair M, Khan MA, et al. Genetic analysis of consanguineous families presenting with congenital ocular defects. Exp Eye Res. 2016;146:163-71.

421. Abouzeid H, Favez T, Schmid A, Agosti C, Youssef M, Marzouk I, et al. Mutations in ALDH1A3 represent a frequent cause of

microphthalmia/anophthalmia in consanguineous families. Hum Mutat. 2014;35(8):949-53.

422. Lin S, Harlalka GV, Hameed A, Reham HM, Yasin M, Muhammad N, et al. Novel mutations in ALDH1A3 associated with autosomal recessive anophthalmia/microphthalmia, and review of the literature. BMC Med Genet. 2018;19(1):160.

423. Semerci CN, Kalay E, Yildirim C, Dincer T, Olmez A, Toraman B, et al. Novel splice-site and missense mutations in the ALDH1A3 gene underlying autosomal recessive anophthalmia/microphthalmia. Br J Ophthalmol. 2014;98(6):832-40.

424. Plaisancie J, Bremond-Gignac D, Demeer B, Gaston V, Verloes A, Fares-Taie L, et al. Incomplete penetrance of biallelic ALDH1A3 mutations. Eur J Med Genet. 2016;59(4):215-8.

425. Yahyavi M, Abouzeid H, Gawdat G, de Preux AS, Xiao T, Bardakjian T, et al. ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. Hum Mol Genet. 2013;22(16):3250-8.

426. Liu Y, Lu Y, Liu S, Liao S. Novel compound heterozygous mutations of ALDH1A3 contribute to anophthalmia in a non-consanguineous Chinese family. Genet Mol Biol. 2017;40(2):430-5.

427. Patel A, Hayward JD, Tailor V, Nyanhete R, Ahlfors H, Gabriel C, et al. The Oculome Panel Test: Next-Generation Sequencing to Diagnose a Diverse Range of Genetic Developmental Eye Disorders. Ophthalmology. 2019;126(6):888-907.

428. Brown R, Hobson RP, Lee A, Stevenson J. Are there “autistic-like” features in congenitally blind children? Journal of Child Psychology and Psychiatry. 1997;38(6):693-703.

429. Brambring M, Asbrock D. Validity of false belief tasks in blind children. J Autism Dev Disord. 2010;40(12):1471-84.

430. Musante L, Ropers HH. Genetics of recessive cognitive disorders. Trends Genet. 2014;30(1):32-9.

431. Lachke SA, Alkuraya FS, Kneeland SC, Ohn T, Aboukhalil A, Howell GR, et al. Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science. 2011;331(6024):1571-6.

432. Alfares A, Alfadhel M, Wani T, Alsahli S, Alluhaydan I, Al Mutairi F, et al. A multicenter clinical exome study in unselected cohorts from a consanguineous population of Saudi Arabia demonstrated a high diagnostic yield. Mol Genet Metab. 2017;121(2):91-5.

433. Maddirevula S, Kuwahara H, Ewida N, Shamseldin HE, Patel N, Alzahrani F, et al. Analysis of transcript-deleterious variants in Mendelian

disorders: implications for RNA-based diagnostics. Genome Biol. 2020;21(1):145.

434. Tan YQ, Tu C, Meng L, Yuan S, Sjaarda C, Luo A, et al. Loss-of-function mutations in TDRD7 lead to a rare novel syndrome combining congenital cataract and nonobstructive azoospermia in humans. Genet Med. 2019;21(5):1209-17.

435. Li D, Wang S, Ye H, Tang Y, Qiu X, Fan Q, et al. Distribution of gene mutations in sporadic congenital cataract in a Han Chinese population. Mol Vis. 2016;22:589-98.

436. Sun Y, Man J, Wan Y, Pan G, Du L, Li L, et al. Targeted next-generation sequencing as a comprehensive test for Mendelian diseases: a cohort diagnostic study. Scientific reports. 2018;8(1):1-9.

437. Yang M, Guo X, Liu X, Shen H, Jia X, Xiao X, et al. Investigation of CYP1B1 mutations in Chinese patients with primary congenital glaucoma. Mol Vis. 2009;15:432-7.

438. Chitsazian F, Tusi BK, Elahi E, Saroei HA, Sanati MH, Yazdani S, et al. CYP1B1 mutation profile of Iranian primary congenital glaucoma patients and associated haplotypes. J Mol Diagn. 2007;9(3):382-93.

439. Tanwar M, Dada T, Sihota R, Das TK, Yadav U, Dada R. Mutation spectrum of CYP1B1 in North Indian congenital glaucoma patients. Mol Vis. 2009;15:1200-9.

440. Keser V, Khan A, Siddiqui S, Lopez I, Ren H, Qamar R, et al. The Genetic Causes of Nonsyndromic Congenital Retinal Detachment: A Genetic and Phenotypic Study of Pakistani Families. Invest Ophthalmol Vis Sci. 2017;58(2):1028-36.

441. Ghiasvand NM, Rudolph DD, Mashayekhi M, Brzezinski JAt, Goldman D, Glaser T. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. Nat Neurosci. 2011;14(5):578-86.

442. Prasov L, Masud T, Khaliq S, Mehdi SQ, Abid A, Oliver ER, et al. ATOH7 mutations cause autosomal recessive persistent hyperplasia of the primary vitreous. Hum Mol Genet. 2012;21(16):3681-94.

443. Khan K, Logan CV, McKibbin M, Sheridan E, Elcioglu NH, Yenice O, et al. Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects. Hum Mol Genet. 2012;21(4):776-83.

444. Atac D, Koller S, Hanson JVM, Feil S, Tiwari A, Bahr A, et al. Atonal homolog 7 (ATOH7) loss-of-function mutations in predominant bilateral optic nerve hypoplasia. Hum Mol Genet. 2020;29(1):132-48.

445. Macgregor S, Hewitt AW, Hysi PG, Ruddle JB, Medland SE, Henders AK, et al. Genome-wide association identifies ATOH7 as a major gene

determining human optic disc size. Human molecular genetics. 2010;19(13):2716-24.

446. Lim SH, St Germain E, Tran-Viet KN, Staffieri S, Marino M, Dollfus PH, et al. Sequencing analysis of the ATOH7 gene in individuals with optic nerve hypoplasia. Ophthalmic Genet. 2014;35(1):1-6.

447. Chen JH, Wang D, Huang C, Zheng Y, Chen H, Pang CP, et al. Interactive effects of ATOH7 and RFTN1 in association with adult-onset primary open-angle glaucoma. Invest Ophthalmol Vis Sci. 2012;53(2):779-85.

448. Ramdas WD, van Koolwijk LM, Lemij HG, Pasutto F, Cree AJ, Thorleifsson G, et al. Common genetic variants associated with open-angle glaucoma. Hum Mol Genet. 2011;20(12):2464-71.

449. Liu Y, Garrett ME, Yaspan BL, Bailey JC, Loomis SJ, Brilliant M, et al. DNA copy number variants of known glaucoma genes in relation to primary open-angle glaucoma. Invest Ophthalmol Vis Sci. 2014;55(12):8251-8.

450. Lewis CJ, Hedberg-Buenz A, DeLuca AP, Stone EM, Alward WLM, Fingert JH. Primary congenital and developmental glaucomas. Hum Mol Genet. 2017;26(R1):R28-R36.

451. Ai M, Heeger S, Bartels CF, Schelling DK, Osteoporosis-Pseudoglioma Collaborative G. Clinical and molecular findings in osteoporosis-pseudoglioma syndrome. Am J Hum Genet. 2005;77(5):741-53.

452. Gilmour DF. Familial exudative vitreoretinopathy and related retinopathies. Eye (Lond). 2015;29(1):1-14.

453. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Human Genetics. 2020:1-11.

454. Landsend ECS, Lagali N, Utheim TP. Congenital Aniridia - A Comprehensive Review of Clinical Features and Therapeutic Approach. Surv Ophthalmol. 2021.

455. Yousaf S, Shahzad M, Kausar T, Sheikh SA, Tariq N, Shabbir AS, et al. Identification and clinical characterization of Hermansky-Pudlak syndrome alleles in the Pakistani population. Pigment Cell Melanoma Res. 2016;29(2):231-5.

456. Fullerton M, McFarland R, Taylor RW, Alston CL. The genetic basis of isolated mitochondrial complex II deficiency. Molecular Genetics and Metabolism. 2020.

457. Zheng C, Wu M, He C-Y, An X-J, Sun M, Chen C-L, et al. RNA granule component TDRD7 gene polymorphisms in a Han Chinese population with age-related cataract. Journal of international medical research. 2014;42(1):153-63.

458. Shiels A, Hejtmancik JF. Molecular Genetics of Cataract. Prog Mol Biol Transl Sci. 2015;134:203-18.

459. Berry V, Georgiou M, Fujinami K, Quinlan R, Moore A, Michaelides M. Inherited cataracts: molecular genetics, clinical features, disease mechanisms and novel therapeutic approaches. British Journal of Ophthalmology. 2020;104(10):1331-7.

460. Al Saai S, McDonald HW, Schey K, Lachke S. Identification of RNA targets and protein binding partners of Tdrd7 in the mouse lens. Investigative Ophthalmology & Visual Science. 2017;58(8):1721-.

461. Shao D-W, Yang C-Y, Liu B, Chen W, Wang H, Ru H-X, et al. Bioinformatics analysis of potential candidates for therapy of TDRD7 deficiency-induced congenital cataract. Ophthalmic research. 2015;54(1):10-7.

462. Barnum CE, Al Saai S, Patel SD, Cheng C, Anand D, Xu X, et al. The Tudor-domain protein TDRD7, mutated in congenital cataract, controls the heat shock protein HSPB1 (HSP27) and lens fiber cell morphology. Human molecular genetics. 2020;29(12):2076-97.

463. Graw J. Mouse models of cataract. Journal of genetics. 2009;88(4):469-86.

464. Richardson R, Tracey-White D, Webster A, Moosajee M. The zebrafish eye—a paradigm for investigating human ocular genetics. Eye. 2017;31(1):68-86.

465. Viet J, Reboutier D, Hardy S, Lachke SA, Paillard L, Gautier-Courteille C. Modeling ocular lens disease in Xenopus. Developmental Dynamics. 2020;249(5):610-21.

466. Hornby SJ, Ward SJ, Gilbert CE, Dandona L, Foster A, Jones RB. Environmental risk factors in congenital malformations of the eye. Ann Trop Paediatr. 2002;22(1):67-77.

467. Kidd JR, Wolf B, Hsia E, Kidd KK. Genetics of propionic acidemia in a Mennonite-Amish kindred. American journal of human genetics. 1980;32(2):236.

468. Schwoerer JS, Candadai SC, Held PK. Long-term outcomes in Amish patients diagnosed with propionic acidemia. Molecular genetics and metabolism reports. 2018;16:36-8.

469. Zollo M, Ahmed M, Ferrucci V, Salpietro V, Asadzadeh F, Carotenuto M, et al. PRUNE is crucial for normal brain development and mutated in microcephaly with neurodevelopmental impairment. Brain. 2017;140(4):940-52.

470. Harlalka GV, McEntagart ME, Gupta N, Skrzypiec AE, Mucha MW, Chioza BA, et al. Novel genetic, clinical, and pathomechanistic insights into

TFG-associated hereditary spastic paraplegia. Human mutation. 2016;37(11):1157-61.

471. Azzariti DR, Hamosh A. Genomic Data Sharing for Novel Mendelian Disease Gene Discovery: The Matchmaker Exchange. Annu Rev Genomics Hum Genet. 2020;21:305-26.

472. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.

473. Raraigh KS, Han ST, Davis E, Evans TA, Pellicore MJ, McCague AF, et al. Functional Assays Are Essential for Interpretation of Missense Variants Associated with Variable Expressivity. Am J Hum Genet. 2018;102(6):1062-77.

474. Marco-Puche G, Lois S, Benitez J, Trivino JC. RNA-Seq Perspectives to Improve Clinical Diagnosis. Front Genet. 2019;10:1152.

475. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 2017;101(3):315-25.

476. Cooper GM. Parlez-vous VUS? Genome Res. 2015;25(10):1423-6.

477. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003;33 Suppl:228-37.

478. Posey JE. Genome sequencing and implications for rare disorders. Orphanet J Rare Dis. 2019;14(1):153.

479. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12(11):745-55.

480. Liu HY, Zhou L, Zheng MY, Huang J, Wan S, Zhu A, et al. Diagnostic and clinical utility of whole genome sequencing in a cohort of undiagnosed Chinese families with rare diseases. Sci Rep. 2019;9(1):19365.

481. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv. 2019:531210.

482. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.

483. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of

inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017;136(6):665-77.

484. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7.

485. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

486. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2-3):377-94.

487. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science. 1987;236(4808):1567-70.

488. Xing C, Schumacher FR, Xing G, Lu Q, Wang T, Elston RC. Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. BMC Genet. 2005;6 Suppl 1:S29.

489. Cooper GM, Mefford HC. Detection of copy number variation using SNP genotyping. Methods Mol Biol. 2011;767:243-52.

490. Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, et al. Pitfalls in homozygosity mapping. Am J Hum Genet. 2000;67(5):1348-51.

491. Benayoun L, Spiegel R, Auslender N, Abbasi AH, Rizel L, Hujeirat Y, et al. Genetic heterogeneity in two consanguineous families segregating early onset retinal degeneration: the pitfalls of homozygosity mapping. Am J Med Genet A. 2009;149a(4):650-6.

492. Lal D, Neubauer BA, Toliat MR, Altmuller J, Thiele H, Nurnberg P, et al. Increased Probability of Co-Occurrence of Two Rare Diseases in Consanguineous Families and Resolution of a Complex Phenotype by Next Generation Sequencing. PLoS One. 2016;11(1):e0146040.

493. Abu-Safieh L, Alrashed M, Anazi S, Alkuraya H, Khan AO, Al-Owain M, et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. Genome Res. 2013;23(2):236-47.

494. Coppieters F, Van Schil K, Bauwens M, Verdin H, De Jaegher A, Syx D, et al. Identity-by-descent-guided mutation analysis and exome sequencing in consanguineous families reveals unusual clinical and molecular findings in retinal dystrophy. Genet Med. 2014;16(9):671-80.

495. Carr IM, Bhaskar S, O'Sullivan J, Aldahmesh MA, Shamseldin HE, Markham AF, et al. Autozygosity mapping with exome sequence data. Hum Mutat. 2013;34(1):50-6.

496. Watson CM, Crinnion LA, Gurgel-Gianetti J, Harrison SM, Daly C, Antanavicuite A, et al. Rapid Detection of Rare Deleterious Variants by Next Generation Sequencing with Optional Microarray SNP Genotype Data. Hum Mutat. 2015;36(9):823-30.

497. Oliveira J, Pereira R, Santos R, Sousa M, editors. Homozygosity mapping using whole-exome sequencing: a valuable approach for pathogenic variant identification in genetic diseases. International Conference on Bioinformatics Models, Methods and Algorithms; 2017: SCITEPRESS.

498. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics. 2018;19(1):106.

499. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 2012;40(10):e72.

500. Ameur A, Kloosterman WP, Hestand MS. Single-Molecule Sequencing: Towards Clinical Applications. Trends Biotechnol. 2019;37(1):72-85.

501. Salzberg SL, Yorke JA. Beware of mis-assembled genomes. Bioinformatics. 21. England2005. p. 4320-1.

502. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133-8.

503. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265-70.

504. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Res. 2017;6:100.

505. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

506. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med. 2017;9(386).

507. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017;8:15824.

508. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012;7(10):e46688.

509. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 1987;15(17):7155-74.

510. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol. 1997;4(3):311-23.

511. Bales KL, Bentley MR, Croyle MJ, Kesterson RA, Yoder BK, Gross AK. BBSome Component BBS5 Is Required for Cone Photoreceptor Protein Trafficking and Outer Segment Maintenance. Investigative Ophthalmology & Visual Science. 2020;61(10):17-.

512. Bentley-Ford MR, Engle SE, Clearman KR, Haycraft CJ, Andersen RS, Croyle MJ, et al. A mouse model of BBS identifies developmental and homeostatic effects of BBS5 mutation and identifies novel pituitary abnormalities. Human Molecular Genetics. 2021.

513. Suspitsin EN, Imyanitov EN. Bardet-Biedl Syndrome. Mol Syndromol. 2016;7(2):62-71.

514. Agha Z, Iqbal Z, Azam M, Hoefsloot LH, van Bokhoven H, Qamar R. A novel homozygous 10 nucleotide deletion in BBS10 causes Bardet-Biedl syndrome in a Pakistani family. Gene. 2013;519(1):177-81.

515. Ajmal M, Khan MI, Neveling K, Tayyab A, Jaffar S, Sadeque A, et al. Exome sequencing identifies a novel and a recurrent BBS1 mutation in Pakistani families with Bardet-Biedl syndrome. Mol Vis. 2013;19:644-53.

516. Ali G, Foo JN, Nasir A, Chang C-H, Chew EG, Latif Z, et al. Identification of a Novel Homozygous Missense (c. 443A> T: p. N148I) Mutation in BBS2 in a Kashmiri Family with Bardet-Biedl Syndrome. BioMed research international. 2021;2021.

517. Billingsley G, Bin J, Fieggen KJ, Duncan JL, Gerth C, Ogata K, et al. Mutations in chaperonin-like BBS genes are a major contributor to disease development in a multiethnic Bardet-Biedl syndrome patient population. J Med Genet. 2010;47(7):453-63.

518. Chen J, Smaoui N, Hammer MB, Jiao X, Riazuddin SA, Harper S, et al. Molecular analysis of Bardet-Biedl syndrome families: report of 21 novel mutations in 10 genes. Invest Ophthalmol Vis Sci. 2011;52(8):5317-24.

519. Deveault C, Billingsley G, Duncan JL, Bin J, Theal R, Vincent A, et al. BBS genotype-phenotype assessment of a multiethnic patient cohort calls for a revision of the disease definition. Hum Mutat. 2011;32(6):610-9.

520. Harville HM, Held S, Diaz-Font A, Davis EE, Diplas BH, Lewis RA, et al. Identification of 11 novel mutations in eight BBS genes by high-resolution homozygosity mapping. J Med Genet. 2010;47(4):262-7.

521. Hayat A, Khan AA, Rauf A, Khan SU, Hussain S, Ullah A, et al. A novel missense variant in the BBS7 gene underlying Bardet-Biedl syndrome in a consanguineous Pakistani family. Clin Dysmorphol. 2020;29(1):17-23.

522. Khan S, Ullah I, Irfanullah, Touseef M, Basit S, Khan MN, et al. Novel homozygous mutations in the genes ARL6 and BBS10 underlying Bardet-Biedl syndrome. Gene. 2013;515(1):84-8.

523. Khan MA, Mohan S, Zubair M, Windpassinger C. Homozygosity mapping identified a novel protein truncating mutation (p.Ser100Leufs*24) of the BBS9 gene in a consanguineous Pakistani family with Bardet Biedl syndrome. BMC Med Genet. 2016;17:10.

524. Maria M, Lamers IJ, Schmidts M, Ajmal M, Jaffar S, Ullah E, et al. Genetic and clinical characterization of Pakistani families with Bardet-Biedl syndrome extends the genetic and phenotypic spectrum. Sci Rep. 2016;6:34764.

525. Muzammal M, Zubair M, Bierbaumer S, Blatterer J, Graf R, Gul A, et al. Exome sequence analysis in consanguineous Pakistani families inheriting Bardet-Biedle syndrome determined founder effect of mutation c.299delC (p.Ser100Leufs*24) in BBS9 gene. Mol Genet Genomic Med. 2019;7(8):e834.

526. Pawlik B, Mir A, Iqbal H, Li Y, Nurnberg G, Becker C, et al. A Novel Familial BBS12 Mutation Associated with a Mild Phenotype: Implications for Clinical and Molecular Diagnostic Strategies. Mol Syndromol. 2010;1(1):27-34.

527. Ullah A, Khalid M, Umair M, Khan SA, Bilal M, Khan S, et al. Novel sequence variants in the MKKS gene cause Bardet-Biedl syndrome with intra- and inter-familial variable phenotypes. Congenit Anom (Kyoto). 2018;58(5):173-5.

528. Ullah A, Umair M, Yousaf M, Khan SA, Nazim-Ud-Din M, Shah K, et al. Sequence variants in four genes underlying Bardet-Biedl syndrome in consanguineous families. Mol Vis. 2017;23:482-94.

529. Forsyth R, Gunay-Aygun M. Bardet-Biedl syndrome overview. GeneReviews®[Internet]: University of Washington, Seattle; 2020.

530. Shah SA, Raheem N, Daud S, Mubeen J, Shaikh AA, Baloch AH, et al. Mutational spectrum of the TYR and SLC45A2 genes in Pakistani families with oculocutaneous albinism, and potential founder effect of missense substitution (p.Arg77Gln) of tyrosinase. Clin Exp Dermatol. 2015;40(7):774-80.

531. Ullah F, Khan I, Shah SA, Azam1+ S, Faisal S, Muzamil H. Mutational Analysis of TYR Gene Causing Oculocutaneous Albinism in Families from District Peshawar. International Journal of Human Genetics. 2019;19(4):170-8.

532. Arshad MW, Harlalka GV, Lin S, D'Atri I, Mehmood S, Shakil M, et al. Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families. Meta Gene. 2018;17:48-55.

533. Kikuchi H, Hara S, Ishiguro S, Tamai M, Watanabe M. Detection of point mutation in the tyrosinase gene of a Japanese albino patient by a direct sequencing of amplified DNA. Human genetics. 1990;85(1):123-4.

534. Kono M, Kondo T, Ito S, Suzuki T, Wakamatsu K, Ito S, et al. Genotype analysis in a patient with oculocutaneous albinism 1 minimal pigment type. British Journal of Dermatology. 2012;166(4):896-8.

535. Tomita Y, Miyamura Y, Kono M, Nakamura R, Matsunaga J. Molecular bases of congenital hypopigmentary disorders in humans and oculocutaneous albinism 1 in Japan. Pigment Cell Res. 2000;13 Suppl 8:130-4.

536. Park KC, Park SK, Lee YS, Youn SW, Park BS, Kim KH, et al. Mutations of the tyrosinase gene in three Korean patients with type I oculocutaneous albinism. Jpn J Hum Genet. 1996;41(3):299-305.

537. Park S-K, Lee K-H, Park K-C, Lee J-S, Spritz RA, Lee S-T. Prevalent and novel mutations of the tyrosinase gene in Korean patients with tyrosinase-deficient oculocutaneous albinism. Molecules & Cells (Springer Science & Business Media BV). 1997;7(2).

538. Khordadpoor Deilamani F, Akbari MT. Potential Founder Effect of Tyrosinase Gene Mutations in Oculocutaneous Albinism Families from West of Iran. Journal of Human Genetics and Genomics. 2017;1(1).

539. Opitz S, Käsmann-Kellner B, Kaufmann M, Schwinger E, Zühlke C. Detection of 53 novel DNA variations within the tyrosinase gene and accumulation of mutations in 17 patients with albinism. Human mutation. 2004;23(6):630-1.

540. Shakil M, Harlalka GV, Ali S, Lin S, D'Atri I, Hussain S, et al. Tyrosinase (TYR) gene sequencing and literature review reveals recurrent mutations and multiple population founder gene mutations as causative of oculocutaneous albinism (OCA) in Pakistani families. Eye. 2019;33(8):1339-46.

541. Ganguly K, Dutta T, Saha A, Sarkar D, Sil A, Ray K, et al. Mapping the TYR gene reveals novel and previously reported variants in Eastern Indian patients highlighting preponderance of the same changes in multiple unrelated ethnicities. Ann Hum Genet. 2020;84(3):303-12.

542. Shah SA, Din SU, Raheem N, Daud S, Mubeen J, Nadeem A, et al. Identification of a novel mutation (p.Ile198Thr) in gene TYR in a Pakistani family with nonsyndromic oculocutaneous albinism. Clin Exp Dermatol. 2014;39(5):646-8.

543. Forshew T, Khaliq S, Tee L, Smith U, Johnson C, Mehdi S, et al. Identification of novel TYR and TYRP1 mutations in oculocutaneous albinism. Clinical genetics. 2005;68(2):182-4.

544. Lu Q, Yuan L, Xu H, Huang X, Yang Z, Yi J, et al. Identification of a missense mutation in the tyrosinase gene in a Chinese family with oculocutaneous albinism type 1. Mol Med Rep. 2017;15(3):1426-30.

545. Rao R, Shu S, Han YZ, Chiu YJ, Han YS. A case report: Co-occurrence of Wilson disease and oculocutaneous albinism in a Chinese patient. Medicine (Baltimore). 2018;97(50):e13744.

546. Zahed L, Zahreddine H, Noureddine B, Rebeiz N, Shakar N, Zalloua P, et al. Molecular basis of oculocutaneous albinism type 1 in Lebanese patients. J Hum Genet. 2005;50(6):317-9.

547. Hutton SM, Spritz RA. Comprehensive analysis of oculocutaneous albinism among non-Hispanic caucasians shows that OCA1 is the most prevalent OCA type. Journal of Investigative Dermatology. 2008;128(10):2442-50.

548. Giebel LB, Tripathi RK, Strunk KM, Hanifin JM, Jackson CE, King RA, et al. Tyrosinase gene mutations associated with type IB ("yellow") oculocutaneous albinism. Am J Hum Genet. 1991;48(6):1159-67.

549. Freudenberg-Hua Y, Freudenberg J, Vacic V, Abhyankar A, Emde AK, Ben-Avraham D, et al. Disease variants in genomes of 44 centenarians. Mol Genet Genomic Med. 2014;2(5):438-50.

550. Rodriguez-Flores JL, Fakhro K, Hackett NR, Salit J, Fuller J, Agosto-Perez F, et al. Exome Sequencing Identifies Potential Risk Variants for M endelian Disorders at High Prevalence in Q atar. Human mutation. 2014;35(1):105-16.

551. Liu N, Kong XD, Shi HR, Wu QH, Jiang M. Tyrosinase gene mutations in the Chinese Han population with OCA1. Genet Res (Camb). 2014;96:e14.

552. S. Oetting W, P. Fryer J, A. King R. A dinucleotide deletion (–ΔGA115) in the tyrosinase gene responsible for type IA (tyrosinase negative) oculocutaneous albinism in a Pakistani individual. Human molecular genetics. 1993;2(7):1047-8.

553. Spritz RA, Fukai K, Holmes SA, Luande J. Frequent intragenic deletion of the P gene in Tanzanian patients with type II oculocutaneous albinism (OCA2). Am J Hum Genet. 1995;56(6):1320-3.

554. Duan HL, Li HY, Wu WQ, Zheng H, Chen Z. [A novel P gene mutation in a Chinese family with oculocutaneous albinism]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2006;23(6):614-7.

555. Kausar T, Jaworek TJ, Tariq N, Sadia S, Ali M, Shaikh RS, et al. Genetic studies of TYRP1 and SLC45A2 in Pakistani patients with nonsyndromic oculocutaneous albinism. J Invest Dermatol. 2013;133(4):1099-102.

556. Rooryck C, Roudaut C, Robine E, Musebeck J, Arveiler B. Oculocutaneous albinism with TYRP1 gene mutations in a Caucasian patient. Pigment Cell Res. 2006;19(3):239-42.

557. Zhang KH, Li Z, Lei J, Pang T, Xu B, Jiang WY, et al. Oculocutaneous albinism type 3 (OCA3): analysis of two novel mutations in TYRP1 gene in two Chinese patients. Cell Biochem Biophys. 2011;61(3):523-9.

558. Saleha SB, Ajaml M, Jamil M, Nasir M, Hameed A. MC1R gene mutation and its association with oculocutaneous albinism type (OCA) phenotype in a consanguineous Pakistani family. Journal of dermatological science. 2013;70(1):68-70.

559. Weisfeld-Adams JD, Mehta L, Rucker JC, Dembitzer FR, Szporn A, Lublin FD, et al. Atypical Chediak-Higashi syndrome with attenuated phenotype: three adult siblings homozygous for a novel LYST deletion and with neurodegenerative disease. Orphanet journal of rare diseases. 2013;8(1):1-7.

560. Yousaf S, Shahzad M, Tasleem K, Sheikh SA, Tariq N, Shabbir AS. Identification and clinical characterization of Hermansky-Pudlak syndrome alleles in the Pakistani population. Pigment cell & melanoma research. 2016;29(2):231.

561. Morgan NV, Pasha S, Johnson CA, Ainsworth JR, Eady RA, Dawood B, et al. A germline mutation in BLOC1S3/reduced pigmentation causes a novel variant of Hermansky-Pudlak syndrome (HPS8). The American Journal of Human Genetics. 2006;78(1):160-6.

562. Badolato R, Prandini A, Caracciolo S, Colombo F, Tabellini G, Giacomelli M, et al. Exome sequencing reveals a pallidin mutation in a Hermansky-Pudlak-like primary immunodeficiency syndrome. Blood. 2012;119(13):3185-7.

APPENDIX A: NGS AND AUTOZYGOSITY MAPPING

A.1 NGS

The traditional approach to disease gene identification in Mendelian diseases involved dideoxy sequencing of candidate genes selected based on their predicted relevance to disease pathophysiology, their location within a disease locus identified through positional mapping, or a combination of both. This approach however is reliant on prior knowledge of the predicted protein structure or function and molecular pathway of disease.

The advent, widespread availability and decreasing costs of high throughput NGS technologies now provide an efficient and cost-effective approach for the unbiased interrogation of an individual's genome. This includes exome sequencing, which investigates the protein-coding regions (exons and splice sites) of the genome, where approximately 85% of Mendelian disease-causing variants are found (477), as well as genome sequencing, which will also allow the detection of variants in non-coding and regulatory genomic regions.

NGS approaches have dramatically accelerated the discovery of novel disease genes for rare Mendelian disorders (478) and have a transformative potential in disease diagnosis and personalised medicine, with genomic medicine initiatives currently being integrated into healthcare systems across the world (86). NGS approaches however often result in large numbers of genomic variants identified [over 20,000 variants in each exome (479) and several million in each genome (480)], and the challenge therefore now shifts from disease gene identification to variant interpretation. This has been aided by the development of population variant databases such as the gnomAD (481), disease variant databases such as ClinVar (482) and HGMD (483), and *in silico* variant pathogenicity prediction tools such as SIFT (484), Polyphen-2 (485) and MaxEntScan (486).

There is however still a need for robust and efficient variant prioritisation strategies to identify the pathogenic variant and prove disease causality. One such strategy

combines a gene sequencing method such as exome or genome sequencing with a genome scanning method such as autozygosity mapping, allowing the rapid identification of potential disease-causing sequence alterations within a disease locus that can be prioritised for further investigation (44).

A.2 Autozygosity mapping

Autozygosity mapping is a powerful technique used to identify candidate loci for autosomal recessive disease genes in populations where there is a significant founder effect due to unions between closely related (consanguineous) or more distantly related individuals (such as endogamous populations with a high frequency of inter-community marriage). This approach relies on the principle that affected individuals in these populations will be homozygous for the same disease variant within a genomic segment or haplotype inherited from a common ancestor, referred to as autozygosity or 'identity by descent' (487). The genome is scanned to identify or 'map' the autozygous regions shared exclusively and consistently by affected individuals, signposting the location where the pathogenic variant is likely to be discovered (Figure A1).

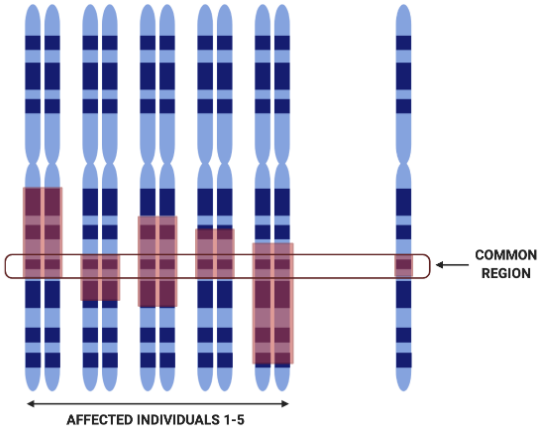


Figure A1 Principles of autozygosity mapping

Red boxes across chromosomes indicate regions of homozygosity in affected individual, with the common autozygous region shared across all affected individuals indicating the most probable location of the disease-causing gene. Image created with BioRender.

Historically, autozygosity mapping was dependent on the analysis of microsatellite markers, however, this has largely been superseded by SNP genotyping panels and microarray technology. SNPs, being diallelic, are less informative than the highly polymorphic microsatellites, but are dispersed more equally throughout the genome

at a higher density, allowing the detection of smaller autozygous regions (488). Additionally, SNP genotyping platforms are easily automated, permitting higher throughput and lower costs, and can also be used to detect CNV (489).

There are important pitfalls to be aware of in the interpretation of autozygosity mapping results, particularly in the study of rare diseases in community settings. Successful identification of disease loci using this technique relies on correct ascertainment of affectation status in family members for accurate inheritance models and interpretation of results. In ophthalmology, this often relies on specialist equipment and expertise that is usually based in the larger teaching hospitals, whilst affected families in communities often reside in more rural and geographically isolated locations with limited accessibility. Additionally, given the enrichment of disease-causing mutations within communities due to a founder effect, a situation may arise where a disease may be caused by compound heterozygous mutations (490, 491), or two phenotypically similar conditions may segregate independently in a single family (especially taking into account the clinical and genetic heterogeneity inherent in many inherited eye diseases) (331, 491, 492), further complicating the interpretation of results generated using autozygosity mapping techniques. Despite these considerations, the combined approach of autozygosity mapping with NGS can helpful in clarifying molecular diagnoses in conditions that are diagnostically challenging due to genetic and phenotypic heterogeneity, such as inherited retinal dystrophies (493, 494).

With the increasingly widespread use of exome and more recently genome sequencing in both diagnostic and research environments, the large volume of sequence data information generated by these NGS technologies raises the possibility of analysing the exome or genome sequence data itself to simultaneously map autozygous regions (without the need for the genome-wide SNP-genotyping microarrays that were previously used for this purpose) and concurrently identify all potentially deleterious sequence variants within these candidate disease loci. This one-step experimental approach is attractive in principle, and several bioinformatic tools have been developed for this purpose (495-497). However, these analyses do have their limitations; with exome data, the coverage can be poor with numerous gaps due to the uneven distribution of coding regions across the genome, and there is the potential for shorter autozygous regions located within gene-poor regions to be missed

(495), whilst the lower read depth of genome sequencing can result in erroneous SNP calls due to errors in sequencing or chance sampling of only one allele at a genuinely heterozygous position (498). With an awareness of these limitations however, adopting a cautious approach to analysis with a greater level of thought and subjective data interpretation, the use of exome or genome data for detecting regions of autozygosity can still be an effective adjunct in the detection of novel disease genes, and for rapid screening for deleterious variants in known disease genes.

A.3 Advances in sequencing technologies

Whilst NGS technologies have facilitated significant advances in medical research, they are not without their limitations. NGS methods are reliant on PCR, which is inefficient in amplifying genomic regions with a high GC content (499). NGS platforms are also dependent on clonal amplification to create clusters of molecules, with the potential for base incorporation errors in individual molecules within a cluster, thereby increasing the noise within samples, and limiting the reads to short lengths of 50-500 bp (500). Numerous short reads require extensive assembly, and genomes often contain repeat sequences that are longer than the NGS reads, resulting in complexities in alignment that can lead to mis-assemblies and gaps (501). Some areas of the genome, such as areas of repetitive DNA, high sequence homology, or even extreme GC content, are therefore still difficult to analyse with current NGS platforms (500).

There are further limitations to the use of NGS technologies for variant analysis in rare diseases. Whilst short reads allow accurate detection of smaller variants such as single-nucleotide variations and short indels, the detection and characterisation of larger structural variants is challenging, and there is limited capacity to link independent variants on the same gene for phasing of alleles and discrimination of parental origin, which is important in situations where compound heterozygosity is identified (233).

The advent of third-generation sequencing, also known as long-read sequencing, may provide a solution to the aforementioned issues. Third-generation sequencing platforms work by single-molecule sequencing in real time with no amplification,

directly providing long linear read lengths (1 – 100 kb) and fast sequencing times (2-10 hours) (500). Two commercially available third-generation sequencing technologies include the single-molecule real-time (SMRT) sequencing by Pacific Biosciences (PacBio) and nanopore sequencing by Oxford Nanopore Technologies (ONT). PacBio utilises a sequencing-by-synthesis approach, with real-time acquisition of signals from fluorescently-labelled nucleotides as they are being incorporated into single DNA molecules (502). Nanopore sequencing instead identifies nucleotides by measuring their electrical conductivity as they pass through the nanopore membrane (503). Both techniques have drawbacks compared to standard NGS platforms, including higher sequencing error rates, lower throughput and higher costs (504), and developments to address these will be needed before these third-generation sequencing technologies can be widely adopted in clinical diagnostic and research settings.

There is also a need for additional biological knowledge that can highlight the molecular signature of the disease. This can aid in the identification of candidate disease genes and variants, particularly for non-coding and regulatory variants, and enable a better understanding of their disease mechanism. One potential route is through functional genomic assays via broad-based ‘omics’ approaches, which study the impact on different cellular products including gene transcripts (transcriptomics), proteins (proteomics) or metabolites (metabolomics), in affected individuals compared to unaffected controls, allowing the identification of aberrant cellular products or activities. One such technique is RNA-sequencing (RNA-seq), which measures transcript levels and diversity in individual tissue types (505), and can be used to highlight the consequences of both coding and non-coding variants on gene expression levels or alternative splicing. Recent applications of this technique have aided in the diagnosis of several rare diseases, with particular success for conditions with a neuromuscular or mitochondrial phenotype (506, 507).

Recent advances in sequencing technologies and bioinformatics have greatly improved our diagnosis and understanding of rare Mendelian disorders. Future efforts are likely to focus on combining the current and developing technologies into a single streamlined workflow for disease variant identification, prioritisation and

characterisation, which will hopefully improve the diagnostic yield in monogenic disorders and shed further light on the complexities of oligogenic diseases.

A.4 Glossary of terms

ClinVar: This is a freely accessible public database that archives and aggregates reports with supporting evidence about relationships between human variation, phenotypes and clinical significance (482).

Genome Aggregation Database (gnomAD): This database was developed by an international coalition of investigators, and aggregates and harmonises both exome and genome sequencing data from a wide variety of large-scale sequencing projects (481). The v2.1.1 dataset (GRCh37) contains 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies. The v3.1.1 dataset (GRCh38) spans 76,156 genomes as selected in v2.

Human Gene Mutation Database (HGMD): This is a database collating germ-line variants in nuclear genes reported in the literature in association with human inherited diseases, and includes both disease-causing variants as well as functional polymorphisms (483).

Minor allele frequency (MAF): the frequency of the alternative allele(s) in a given population.

Mapping quality (MQ): This is the root mean square of the mapping quality of reads across all reads at the site, and provides an estimation of the overall mapping quality of reads supporting a variant call.

MaxEntScan: This *in silico* algorithm is based on the ‘maximum entropy principle’ (486). The splice signal given by a wild-type sequence is compared to the splice site signal given by a mutated sequence, and a change in splice site signal of ≥10% is considered to predict a pathogenic effect.

Polymorphism Phenotyping v2 (Polyphen-2): This *in silico* algorithm applies to non-synonymous variants, and predicts the functional significance of an amino acid substitution using sequence-based and structure-based predictive features (485). Scores range from 0.0 (tolerated) to 1.0 (deleterious).

Protein Variation Effect Analyser (PROVEAN): This *in silico* algorithm applies to non-synonymous and indel variants, and predicts whether an amino acid substitution or indel has an impact on the biological function of a protein based on the degree of conservation of amino acid residues in sequence alignments derived from closely-related sequences (508). If the generated score is ≤ -2.5, the protein variant is predicted to have a deleterious effect; if the score is above the threshold, the variant is predicted to have a neutral effect.

Read depth (DP): This is the number of reads that have passed the variant caller's internal quality control metrics.

Sorting Intolerant From Tolerant (SIFT): This *in silico* algorithm applies to non-synonymous variants, and predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids (484). Scores range from 0.0 (deleterious) to 1.0 (tolerated).

SpliceSiteFinder-like (SSF): This prediction program uses an algorithm based on Shapiro and Senapathy *et al* (509) using position weight matrices computed from a set of human constitutive exon/intron junctions for donor and acceptor sites. The splice signal given by the wild-type sequence is compared to the splice site signal given by the mutated sequence. A change in splice site signal of ≥10% is considered to predict a pathogenic effect.

Splice Site Prediction by Neural Network (NNSplice): This prediction algorithm is based on neural networks, and analyses the structure of donor and acceptor sites using a separate neural network recogniser for each site (510).

APPENDIX B: A NOVEL *BBS5* VARIANT ASSOCIATED WITH BBS IN A PAKISTANI FAMILY

Clinical and genomic investigations were undertaken in a Pakistani family (family 42) where affected individuals displayed phenotypic features that were highly suggestive of BBS, in order to aid diagnosis and clinical management.

B.1 Materials and methods

Family recruitment, clinical assessment, blood sample collection and DNA extraction was performed as previously described (see section 4.2.2). WES was undertaken using DNA from a single affected individual in family 42 (individual IV:1) at BGI Hong Kong, as described in section 2.3.5. Bioinformatic analysis with additional virtual gene panel analysis and filtering to retain heterozygous variants compatible with triallelism was performed as described in section 4.2.2. Primer design, PCR and dideoxy sequencing (Appendix Table D2) was also performed as previously described in section 2.3.3 to genotype and confirm appropriate segregation of the candidate disease variant in all available affected and unaffected individuals.

A literature review was performed as described in section 2.4 to retrieve all BBS-associated variants reported in Pakistani families. Findings are summarised in Table B2 and Figure B2.

B.2 Results: clinical and genetic findings

A multigenerational family (family 42) with two affected individuals (IV:1 and IV:4) residing in a remote village in the KPK province of Pakistan was investigated. Both affected individuals were described to have moderate developmental delay/intellectual disability, visual impairment, truncal obesity, postaxial polydactyly, and renal anomalies. A single affected male was also reported to have hypogonadism (individual IV:1) (Table B1).

Table B1 Summary of clinical features observed in affected individuals in family 42 with BBS and homozygous for the *BBS5* c.196delA variant

Individual	Gender	Age (yrs)	ID	Obesity	Visual Impairment	Hypogonadism	Renal impairment	Post-axial polydactyly
IV:1	M	32	✓ (S)	✓	✓ Progressive loss of vision (S), photosensitivity, poor night vision Retinal dystrophy and optic disc pallor	✓	✓	✓
IV:4	F	8	✓ (S)	✓	✓ Progressive loss of vision (M) Optic disc pallor	✗	✓	✓

Abbreviations: F, female; M, male; yrs, years; (S), severe/profound; (M), mild/moderate. The (✓) and (✗) symbols indicate the presence of absence of a feature in an affected subject respectively.

After filtering of variants for zygosity, call quality, population frequency and predicted outcome in WES data in a single affected individual (IV:1), only a single plausible candidate disease variant of potential relevance to the phenotype was identified, a homozygous novel *BBS5* frameshift variant (GRCh38) chr2:g.169487122delA; NM_152384.2:c.196delA; p.(Arg66Glufs*12). This variant was absent in gnomAD (v2.1.1 and v3.1.1) and from a control exome dataset of 100 ethnically matched controls undertaken in the Human Molecular Genetics laboratory in Pakistan. This variant segregated appropriately for an autosomal recessive condition in the family (Figure B1).

Additional analysis of exome data undertaken using the “rare multisystem ciliopathy disorders v1.84” PanelApp virtual gene panel did not identify any homozygous or compound heterozygous candidate variants compatible with the phenotype. No heterozygous variants compatible with triallelism were identified.

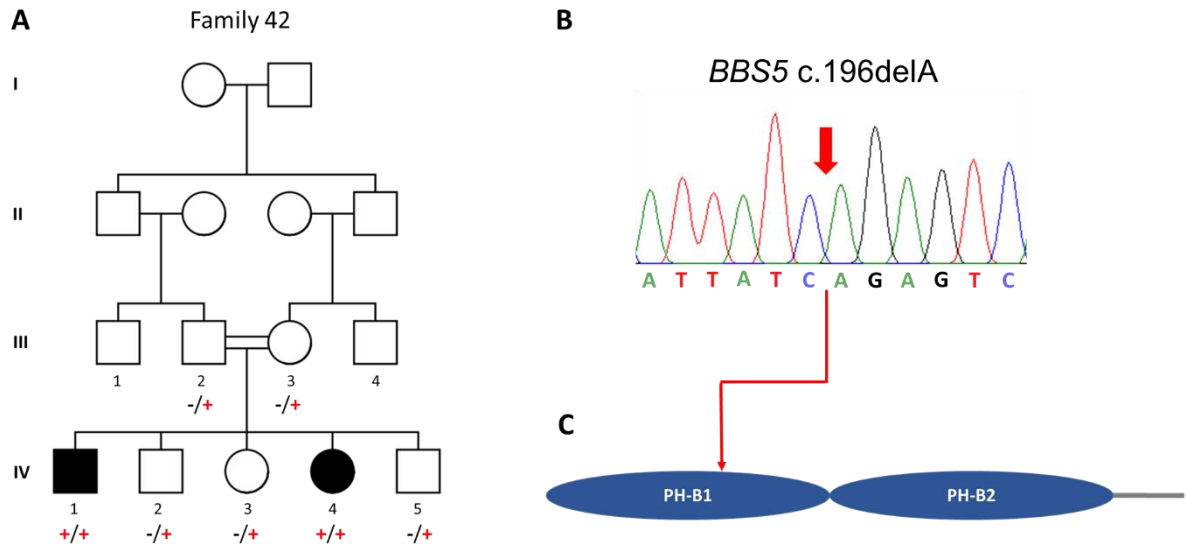


Figure B1 Family 42 pedigree showing *BBS5* c.196delA genotype data

(A) Pedigree diagram of family 42 showing segregation of the BBS5 *c.196delA; p.(Arg66Glufs*12) variant, co-segregating appropriately for an autosomal recessive condition.. Genotypes are shown beneath generations III and IV (+, c.196delA; -, wild type). (B) Sequence chromatogram of* BBS5 *c.196delA in a homozygous affected individual. (C) Schematic showing domain architecture of BBS5 [adapted from Nachury* et al (287) *and the location of the c.196delA; p.(Arg66Glufs*12) variant. This variant is located within the first of two predicted Pleckstrin Homology-like (PH-like) domains (PH-B1 and PH-B2).*

B.3 Discussion

BBS5, located on chromosome 2q31.1, encodes the BBS5 protein, one of eight proteins that form the stable core of the BBSome protein complex. BBS5 is structurally and functionally unique among the BBSome components in that it has two Pleckstrin Homology-like (PH-like) domains that are capable of binding to phosphoinositides (Figure B1), with this interaction predicted to play a critical role in ciliogenesis (287). *Bbs5* knockout or mutant mice develop a complex phenotype consisting of increased pre-weaning lethality, craniofacial and skeletal defects, ventriculomegaly, pituitary anomalies, obesity, retinal degeneration and male infertility; recapitulating features previous described in ciliopathy syndromes (511, 512).

Here, WES analysis identified a novel homozygous variant in *BBS5*, confirming a diagnosis of BBS as likely underlying the clinical features in the affected family (family 42) investigated. This *BBS5* c.196delA variant is predicted to result in a frameshift [p.(Arg66fsGlufs*12)] and nonsense-mediated mRNA decay. The two affected individuals demonstrate classical features of BBS, including postaxial polydactyly,

learning difficulties, retinitis pigmentosa, and truncal obesity (284). Intra-familial phenotypic variability was noted, with individual IV:1 displaying more significant visual impairment and intellectual disability compared to his affected sister. Bilateral postaxial polydactyly of the lower limbs and clinodactyly of the hands was identified on examination in individual IV:1, whereas his sister IV:4 had bilateral postaxial polydactyly in both upper and lower limbs and short toes. This intra-familial phenotypic variability is well described in BBS, and "triallelism" has been proposed as a possible mechanism for this, where affected individuals carry homozygous or compound heterozygous variants at one BBS locus with a third heterozygous variant in a second BBS gene that interacts to cause or modify the disease phenotype (254). However, virtual gene panel analysis of BBS genes undertaken in the present study did not identify any candidate variants that could plausibly be compatible with triallelism.

In populations of Northern European descent, *BBS1* and *BBS10* are the most commonly mutated genes, responsible for approximately 23% and 20% of BBS cases respectively (300, 513), with the *BBS1* p.(Met390Arg) (297) and *BBS10* p.(Cys91fsLeufs*5) variants (307) being the most common pathogenic alleles for each of the two genes. In Pakistan, variants in ten out of the 21 BBS-associated genes have been identified in 74 affected individuals from 28 BBS families (Table B2) (61, 80, 511, 514-528), including two individuals from the current study family (Figure B2). Although *ARL6* and *TTC8* are minor contributors to BBS globally, within the Pakistani community, they account 13.3% and 12.0% of all BBS cases respectively, and are the two most commonly mutated genes associated with BBS in Pakistan (Figure B2).

To date, there have only been two *BBS5* variants associated with BBS in Pakistani families: a homozygous missense variant of the start codon [c.2T>A; p.(M1?)] likely resulting in either failure of protein translation or in the translation of an illegitimate transcript reported in two families (520), and a 11-bp deletion [c.743_744del11; p.(Glu245Glyfs*18)] frameshift mutation likely resulting in loss of function in a further two families (524) (Table B2). The novel *BBS5* variant identified in this study therefore significantly expands on the contribution of *BBS5* mutations to BBS in Pakistan.

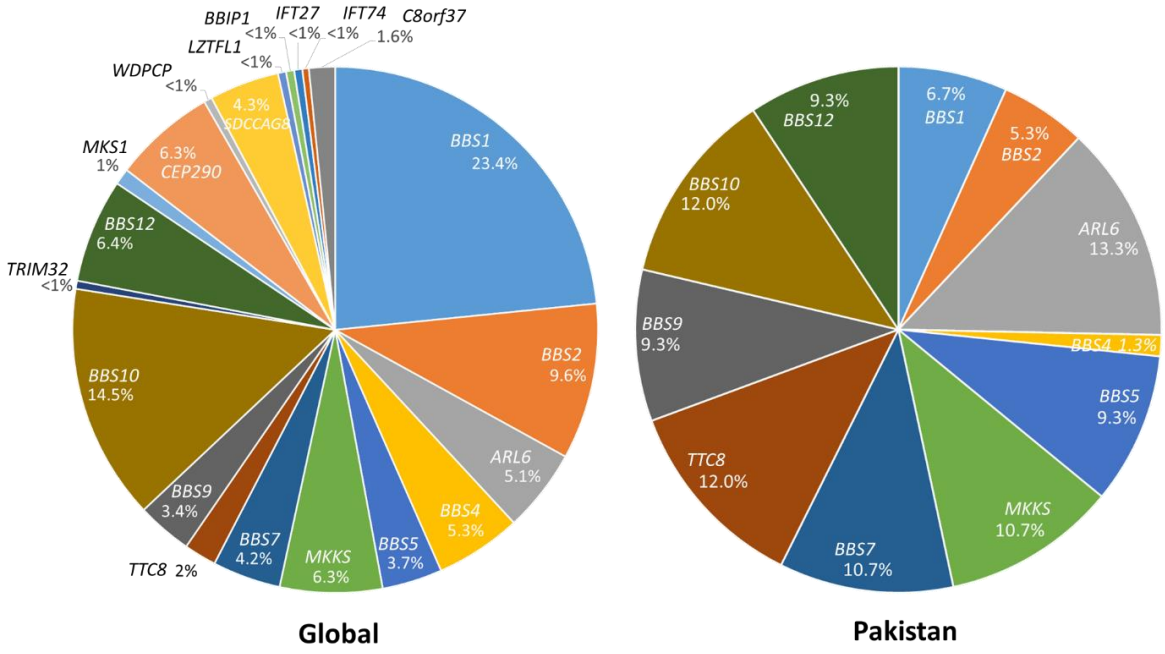


Figure B2 Contribution of BBS genes to BBS globally and within Pakistani families

Global figures derived from (529)

Table B2 Summary of all reported variants associated with BBS in Pakistan

Gene	Nucleotide variant	Protein variant	Number of reported families (individuals)	Region in Pakistan	Reference	ClinVar (Accession)
BBS1 (NM_024649.5)	c.47+1G>T	Affects splicing	1 (2)	Central Punjab	Ajmal *et al* (515)	Not present
	c.442G>A	p.(Asp148Asn)	1 (2)	Central Punjab	Ajmal *et al* (515)	Likely pathogenic (VCV000645579)
	c.887delT	p.(Ile296Thrfs*7)	1 (1)	NA	Billingsley *et al* (517)	Pathogenic (VCV000193740)
BBS2 (NM_031885.5)	c.443A>T	p.(Asn148Ile)	1 (2)	Bagh, AJ&K	Ali *et al* (516)	Not present
	c.1237C>T	p.(Arg413*)	2 (2)	NA	Harville *et al* (520); Chen *et al* (518)	Pathogenic/ likely pathogenic (VCV000370943)
ARL6 (NM_177976.3)	c.123+1131_*25296del53985	53985bp deletion	1 (1)	NA	Chen *et al* (518)	Not present
	c.281T>C	p.(Ile94Thr)	1 (7)	Remote regions	Khan *et al* (522)	Likely pathogenic
	c.534A>G	p.(Gln178Gln)	1 (2)	NA	Maria *et al* (524)	Not present
BBS4 NM_033028.5	c.221-1G>A	Affects splicing	1 (2)	Quetta, Balochistan	Personal communication – Dr I D'Atri	Not present
BBS5 (NM_152384.3)	c.2T>A	p.(Met1Lys)	2 (2)	NA	Harville *et al* (520)	Pathogenic (VCV000812118)
	c.196delA	p.(Arg66Glufs*12)	1 (2)	Bannu, KPK	Khan *et al* (61) – this study	Not present
	c.734_744del11	p.(Glu245Glyfs*18)	2 (3)	NA	Maria *et al* (524)	Not present
MKKS (NM_018848.3)	c.119C>G	p.(Ser40*)	1 (3)	Nawab Shah City, Sindh	Ullah *et al* (a) (528)	Likely pathogenic (VCV000549478)
	c.287C>T	p.(Ala96Val)	1 (3)	NA	Ullah *et al* (b) (527)	Not present

	c.775delA	p.(Thr259Leufs*21)	1 (2)	NA	Ullah *et al* (b) (527)	Pathogenic (VCV000866319)
BBS7 (NM_176824.3)	c.580_582delGCA	p.(Ala194del)	1 (3)	KPK	Ullah *et al* (a) (528)	Not present
	c.719G>T	p.(Gly240Val)	1 (2)	Peshawar, KPK	Hayat *et al* (521)	Not present
	c.1592_1597delTTCCAG	p.(Val531_Pro532del)	1 (3)	AJ&K	Ullah *et al* (a) (528)	Pathogenic (VCV000030680)
TTC8 (NM_198309.3)	c.235+1G>A	Affects splicing	1 (2)	NA	Deveault *et al* (519)	Not present
	c.1019+2_1019+4del	Affects splicing	2 (4)	NA	Harville *et al* (520); Ansley *et al* (80)	Pathogenic (VCV000002529)
	c.1347G>C	p.(Gln449His)	1 (3)	KPK	Ullah *et al* (a)(528)	Pathogenic (VCV000235131)
BBS9 (NM_198428.3)	c.299delC	p.(Ser100Leufs*24)	3 (6)	DI Khan, KPK	Khan *et al* (523); Muzammal *et al* (525)	Not present
	c.1789C>T	p.(Gln597*)	1 (1)	NA	Maria *et al* (524)	Not present
BBS10 (NM_024685.4)	c.271dupT	p.(Cys91Leufs*5)	1 (2)	AJ&K	Ullah *et al* (a) (528)	Pathogenic (VCV000001328)
	c.1075C>T	p.(Gln359*)	1 (5)	Remote regions	Khan *et al* (522)	Not present
	c.1958_1967del	p.(Ser653Ilefs*4)	1 (2)	Punjab	Agha et al (514)	Not present
BBS12 (NM_152618.3)	c.1438delG	p.(Arg480Metfs*3)	1 (3)	NA	Billingsley *et al* (517)	Not present
	c.1589T>C	p.(Leu530Pro)	2 (2)	NA	Harville *et al* (520)	Not present
	c.2102C>A	p.(Ser701*)	1 (3)	NA	Pawlick *et al* (526)	Not present

Abbreviations: AJ&K, Azad Jammu and Kashmir; BBS, Bardet-Biedl syndrome; KPK, Khyber Pakhtunkhwa; NA, not available. The novel BBS5 variant identified in this study is highlighted in yellow

APPENDIX C: GENETIC VARIANTS AND MAPPED LOCI ASSOCIATED WITH NON-SYNDROMIC AND SYNDROMIC OCA IN PAKISTANI POPULATIONS

Appendix Table C includes a summary of ClinVar entries and PubMed reports defining possible pathogenicity of each variant, as well as other genetic information including the number of families and affected individuals reported

Gene	Nucleotide variant	Protein variant	References	Province (region) [caste]	No of affected individuals (families)	Variant identified elsewhere	ClinVar (Accession)
TYR	c.62C>T	p.(Pro21Leu)	Jaworek *et al* (193)	Punjab [Sayyid]	1 (15)	No	Not present
TYR	c.103T>C	p.(Cys35Arg)	Jaworek *et al* (193)	Punjab [Malik; Malik Jutt]	2 (10)	No	Not present
TYR	c.132T>A	p.(Ser44Arg)	Shah[a] *et al* (530); Ullah *et al* (531); Arshad *et al* (532); Unpublished	KPK (Peshawar)	4 (13)	No	Not present
TYR	c.164G>C	p.(Cys55Ser)	Shahzad *et al* (186)	Punjab [Sidhu jut]	1 (3)	No	VUS (VCV000617794)
TYR	c.223G>T	p.(Asp75Tyr)	Shahzad *et al* (186)	Punjab [Qureshi]	1 (3)	No	Likely pathogenic (VCV000617795)
TYR	c.230G>A	p.(Arg77Gln)	Shah[a] *et al* (530)	AJ&K [Khokar; Rajput]	5 (20)	Japanese (201, 223, 226, 533-535); Korean (536, 537); Chinese (162, 189, 229); Iran (538); Caucasian (129); North European (131); Scottish (131); Italian (131); German (539); France (153)	Pathogenic/ likely pathogenic (VCV000003776)
TYR	c.240G>C	p.(Trp80Cys)	Arshad *et al* (532)	KPK	1 (2)	No	Not present

				[Pathan]			
TYR	c.248T>G	p.(Val83Gly)	Gul[a] *et al* (188)	KPK (Swat) [Pashto]	1 (1)	No	Not present
TYR	c.272G>A	p.(Cys91Tyr)	Shakil *et al* (540)	Punjab (RYK) [Somro]	1 (4)	North American Hutterites (195)	Pathogenic (VCV000039977)
TYR	c.308G>A	p.(Cys103Tyr)	Shakil *et al* (540)	Punjab (Sahiwal) [Khokar]	1 (4)	France (141)	Not present
TYR	c.346C>T	p.(Arg116*)	Gul[b] *et al* (192); Shakil *et al* (540)	Punjab (Lahore) [Pukhtun; Mughal]	2 (3)	Chinese (131, 162, 189, 197, 229); German (539)	Pathogenic (VCV000099565)
TYR	c.585G>A	p.(Trp195*)	Shahzad *et al* (186)	Punjab [Gondal]	1 (6)	Indian (541)	Likely pathogenic (VCV000617796)
TYR	c.593T>C	p.(Ile198Thr)	Shah[b] *et al* (542)	NA	1 (4)	No	Not present
TYR	c.649C>T	p.(Arg217Trp)	Shahzad *et al* (186); Sajid *et al* (191); Arshad *et al* (532)	Punjab; Sindh [Rajput; Alvi; Mochi; Arain; Minhas]	8 (41)	Caucasian (198)	Pathogenic/ likely pathogenic/ VUS (VCV000003795)
TYR	c.667C>T	p.(Gln223*)	Unpublished	KPK (Bunar) [Pashton]	1 (2)	No	Not present
TYR	c.715C>T	p.(Arg239Trp)	Shah[a] *et al* (530); Shakil *et al* (540)	Punjab (RYK) [Turk]	2 (6)	Japanese (199); South Indian (200); Chinese (189)	Not present
TYR	c.826T>C	p.(Cys276Arg)	Bibi *et al* (205)	KPK	1 (3)	No	Not present
TYR	c.832C>T	p.(Arg278*)	Shahzad *et al* (186); Jaworek *et al* (193); Bibi *et al* (205); Sajid *et al* (191); Forshew *et al* (543); Arshad *et al* (532); Shakil *et al* (540); Unpublished	Punjab; Sindh; KPK (Lahore; Peshawar) [Punjabi; Shaikh; Rajput; Mula khel; Arain; Malik; Jutt; Sayyid bukhari; Bhatti; Urdu-speaking; Gujjar; Mayo; Pashton]	22 (99)	Guyana (194); Japanese (201, 226, 227); Chinese (162, 171, 189, 197, 202, 228-230); Korean (231); Indian (131, 215, 223-225); Canada (203); Irish (131); Italian (131); Mexican (131); South European (131); Syrian (131); German	Pathogenic (VCV000099583)

						(131); Polish (131); Israeli Moroccan/Tunisian Jewish (232)	
TYR	c.895C>T	p.(Arg299Cys)	Gul[a] *et al* (188)	Punjab (Ranwal) [Saraiki]	1 (2)	Danish (146); Chinese (197);	Not present
TYR	c.896G>A	p.(Arg299His)	Shahzad *et al* (186); Jaworek *et al* (193); Mauri *et al* (140)	Punjab [Warraich; Abbassi; Waseer]	4 (15)	Caucasian (198); Chinese (189, 197, 230, 544, 545); Israeli Arab Christian (232); Korean (537)	Pathogenic (VCV000003796)
TYR	c.982G>C	p.(Glu328Gln)	Tripathi *et al* (194)	NA	1 (1)	No	Not provided (VCV000099591)
TYR	c.1037G>T	p.(Gly346Val)	Sajid *et al* (191)	Sindh	1 (4)	Indian (223)	Likely pathogenic (VCV000437986)
TYR	c.1037G>A	p.(Gly346Glu)	Shahzad *et al* (186)	Punjab [Sahoo]	1 (3)	Chinese (189); Lebanese (546)	Likely pathogenic (VCV000617799)
TYR	c.1147G>A	p.(Asp383Asn)	Shahzad *et al* (186)	Punjab [Bangulzai]	1 (4)	Japanese (534); North European (131); French (131); German (131, 539); Norwegian (131); Irish (131); Caucasian (134, 547); Korean (536)	Pathogenic (VCV000003775)
TYR	c.1204C>T	p.(Arg402*)	Shahzad *et al* (186)	Punjab [Gujjar]	1 (5)	Israeli Arab Christian (232); East Indian (215); Chinese (189); France (153); English (131); Scottish (131); German (131, 539); Italian (131); Lebanese (131, 546); Hungarian (131); Irish (131); Polish (131); Caucasian (134, 547)	Pathogenic (VCV000099542)

TYR	c.1217C>T	p.(Pro406Leu)	Jaworek *et al* (193)	Punjab [Malik Jutt]	1 (6)	France (153); Amish (548); Ashkenazi Jewish (549); Qatari (550)	Pathogenic (VCV000003777)
TYR	c.1231T>C	p.(Tyr411His)	Jaworek *et al* (193)	Punjab [Arain]	1 (6)	No	Not present
TYR	c.1255G>A	p.(Gly419Arg)	Shahzad *et al* (186); Gul[a] *et al* (188); Jaworek *et al* (193); Gul[b] *et al* (192); Tripathi *et al* (194); Sajid *et al* (191); King *et al* (131); Arshad *et al* (532); Shakil *et al* (540); Unpublished	Punjab; KPK; Sindh; AJ&K (Chiniot; Gujranwala; Lahore; Sargodha; Peshawar; DI Khan; Swat) [Virk; Saraiki; Pashto; Bhatti Jutt; Rajput; Shaikh; Chanarr; Daaye; Oteera; Dodiyaanay sial; Sindhi speaking; Urdu speaking; Saraiki speaking; Chudhar Jutt; Arain; Malik Awan; Cheema; Pashton]	26 (94)	Indian (207, 215, 224); Egyptian (131); German (539); Caucasian (547)	Pathogenic (VCV000003792)
TYR	c.1424G>A	p.(Trp475*)	Gul[a] *et al* (188)	Punjab (Bhakkar) [Saraiki]	1 (5)	No	Not present
TYR	c.1037-18T>G	Affects splicing	Shahzad *et al* (186)	Sindh [Urdu-speaking]	1 (4)	No	Likely pathogenic (VCV000617798)
TYR	c.1037-7T>A	Affects splicing	Shahzad *et al* (186); Gul[a] *et al* (188); Unpublished	Punjab; KPK (Sargodha; DI Khan)	6 (22)	Chinese (189); France (153); Danish (131); North European (131); Italian (131); Spanish	Pathogenic/ likely pathogenic (VCV000099527)

				[Saraiki; Goraya; Sayyid bukhari; Khokar; Cheema]		(131); German (131); Caucasian (134, 547); Israeli Moroccan/Sephardic Jewish (232)	
TYR	c.1184+2T>C	Affects splicing	Shahzad *et al* (186)	Punjab (Bhatti)	1 (6)	Chinese (551)	Likely pathogenic (VCV000617800)
TYR	c.344_345del GA	p.(Arg115Thrfs*52)	Oetting *et al* (552)	NA	1 (1)	German (539); North European (131); Caucasian (547)	Not present
TYR	c.943_948del TCAGCT	p.(Ser315_Ala316 delSerAla)	Shahzad *et al* (186); Forshew et al (543)	Punjab [Butt]	2 (14)	No	Likely pathogenic (VCV000617797)
TYR	c.1002delA	p.(Ala335Leufs*20)	Unpublished	Punjab (Lahore) [Mayo]	1 (3)	No	Not present
TYR	Exon 4-5 del	Frameshift	Shahzad *et al* (186)	Punjab [Jutt; Chakotaray]	2 (23)	South Indian (Vysya) (224)	-
OCA2	c.827T>A	p.(Val276Glu)	Sajid *et al* (191)	AJ&K	1 (2)	No	Not present
OCA2	c.877G>C	p.(Glu293Gln)	Sajid *et al* (191)	AJ&K	1 (2)	No	Not present
OCA2	c.954G>A	p.(Met318Ile)	Jaworek *et al* (193)	Punjab [Warraich]	1 (23)	No	Not present
OCA2	c.1056A>C	p.(Arg352Ser)	Shahzad *et al* (186)	Sindh [Sindhi-speaking]	1 (11)	No	Not present
OCA2	c.1064C>T	p.(Ala355Val)	Shahzad *et al* (186)	Punjab [Chandia]	1 (5)	No	Likely pathogenic (VCV000617804)
OCA2	c.1075G>C	p.(Gly359Arg)	Shahzad *et al* (186)	KPK [Saraiki-speaking]	1 (11)	No	Likely pathogenic (VCV000452941)
OCA2	c.1211C>T	p.(Thr404Met)	Shahzad *et al* (186)	Punjab [Sayyid]	1 (5)	Nigerian (553)	Likely pathogenic (VCV000211766)
OCA2	c.1322A>G	p.(Asp441Gly)	Shahzad *et al* (186)	Sindh [Sindhi-speaking]	1 (4)	No	VUS (VCV000502704)
OCA2	c.1327G>A	p.(Val443Ile)	Mauri *et al* (140); Arshad *et al* (532)	Punjab [Saraiki]	2 (2)	France (153); North European (212); German (217); Chinese (211)	Pathogenic/ likely pathogenic/ VUS (VCV000000955)

OCA2	c.1456G>T	p.(Asp486Tyr)	Shahzad *et al* (186); Jaworek *et al* (193); Sajid *et al* (191); Unpublished	Punjab; KPK; Sindh (Qasur; Borewala) [Lanjay; Mehay; Ghallu; Chaaki; Wains; Bhutta; Saraiki-speaking; Sindhi-speaking; Dogar]	12 (61)	No	VUS (VCV000617806)
OCA2	c.1580T>G	p.(Leu527Arg)	Jaworek *et al* (193);	Punjab [Warraich]	1 (23)	No	Not present
OCA2	c.1762C>T	p.(Arg588Trp)	Arshad *et al* (532)	Punjab [Saraiki]	1 (1)	No	VUS (VCV000885235)
OCA2	c.1922C>T	p.(Ser641Leu)	Shahzad *et al* (186)	Punjab [Mughal]	1 (4)	No	Likely pathogenic (VCV000617807)
OCA2	c.2020C>G	p.(Leu674Val)	Arshad *et al* (532)	KPK [Yousafzai]	1 (2)	East Indian (215)	Pathogenic/ VUS (VCV000194918)
OCA2	c.2207C>T	p.(Ser736Leu)	Unpublished	NA	1 (1)	Caucasian (208); France (153); China (209)	Likely pathogenic/ VUS (VCV000195557)
OCA2	c.2228C>T	p.(Pro743Leu)	Shahzad *et al* (186); Jaworek *et al* (193)	Punjab; Sindh [Arain; Joyia; Urdu-speaking]	3 (19)	North European (212); Chinese (189)	Pathogenic/ likely pathogenic (VCV000000956)
OCA2	c.2359G>A	p.(Ala787Thr)	Jaworek *et al* (193)	Punjab [Chohan]	1 (4)	France (153); Chinese (554)	Pathogenic (VCV000803060)
OCA2	c.2360C>T	p.(Ala787Val)	Shahzad *et al* (186); Unpublished	KPK, Punjab (Bajaur) [Arain]	2 (6)	Yes (210)	Pathogenic/ likely pathogenic (VCV000617810)
OCA2	c.2360C>A	p.(Ala787Glu)	Shahzad *et al* (186)	Punjab [Langah Jutt]	1 (9)	No	Likely pathogenic (VCV000617809)
OCA2	c.2458T>C	p.(Ser820Pro)	Arshad *et al* (532)	Punjab [Niaz]	1 (2)	No	Not present
OCA2	c.1045-15T>G	Affects splicing	Shahzad *et al* (186); Jaworek *et al*	Punjab; KPK; Sindh (Peshawar) [Niaz; Afridi;	18 (70)	No	Likely pathogenic (VCV000617802)

			(193); Arshad *et al* (532); Unpublished	Bubar; Sidhu Jutt; Abbassi; Ansari; Rajput; Jutt; Arain, Bhatti; Sindhi-speaking; Saraiki-speaking; Pashton]			
OCA2	c.1182+2dupT	Affects splicing	Shahzad *et al* (186)	Punjab [Arain]	1 (3)	No	Likely pathogenic (VCV000617805)
OCA2	c.1951+4A>G	Affects splicing	Shahzad *et al* (186)	Sindh [Urdu-speaking]	1 (5)	No	Likely pathogenic (VCV000617808)
OCA2	c.408_409delAA	p.(Arg137Ilefs*83)	Arshad *et al* (532); Unpublished	KPK (Peshawar) [Yousafzai; Pashton]	2 (4)	No	Not present
OCA2	c.987delCinsAGA	p.(Gln330Aspfs*2)	Shahzad *et al* (186)	Sindh	1 (5)	No	Likely pathogenic (VCV000617801)
OCA2	c.1960delG	p.(Ala654Leufs*8)	Lee *et al* (212)	NA	1 (3)	No	Pathogenic (VCV000000957)
OCA2	Exon 3-14 del	Frameshift	Shahzad *et al* (186)	Punjab [Machi]	1 (3)	No	-
OCA2	Exon 7-8 del	Frameshift	Shahzad *et al* (186); Mauri *et al* (140)	Punjab [Mughal; Terkhan]	2 (8)	No	-
OCA2	Exon 19 del	Frameshift	Shahzad *et al* (186); Gul[a] *et al* (188); Unpublished	Punjab (Tank) [Pashto; Malik]	4 (18)	No	-
OCA2	Exon 20-24 del	Frameshift	Shahzad *et al* (186)	Punjab [Achakzai]	1 (5)	No	-
OCA2	Exon 22-24 del	Frameshift	Shahzad *et al* (186)	Punjab [Rajput]	1 (3)	No	-
TYRP1	c.256G>T	p.(Asp86Tyr)	Kausar[a] *et al* (555)	Punjab [Bhutta]	1 (3)	No	Likely pathogenic (VCV000617811)
TYRP1	c.1067G>A	p.(Arg356Gln)	Shahzad *et al* (186)	Punjab [Langah; Haraaj; Matam]	3 (12)	Caucasian (556), Chinese (557)	Pathogenic/ likely pathogenic (VCV000017596)

TYRP1	c.1120C>T	p.(Arg374*)	Forshew *et al* (543)	NA	1 (8)	No	Pathogenic (VCV000017595)
TYRP1	c.1534C>T	p.(Glu512*)	Shahzad *et al* (186)	Punjab [Shaikh]	1 (7)	No	Likely pathogenic (VCV000617812)
TYRP1	c.647_668del	p.(Glu216Glyfs*42)	Kausar[a] *et al* (555)	Punjab [Ladd]	1 (5)	No	Not present
TYRP1	~1 Mb del including whole gene		Gul[a] *et al* (188)	KPK (Lakki Marwat) [Pashto]	1 (2)	No	-
SLC45A2	c.251T>C	p.(Leu84Pro)	Kausar[a] *et al* (555)	Punjab [Lakhat]	1 (4)	No	Not present
SLC45A2	c.1532C>T	p.(Ala511Val)	Shahzad *et al* (186); Gul[a] *et al* (188); Kausar[a] *et al* (555)	Punjab; KPK (Kech) [Saraiki; Saraiki-speaking; Kingray]	4 (17)	No	Likely pathogenic (VCV000617813)
SLC45A2	c.889-6T>G	Affects splicing	Kausar[a] *et al* (555)	Punjab [Rehmani]	1 (3)	No	Not present
SLC45A2	c.1331_dupA	p.(Asn444Lysfs*5)	Bibi *et al* (205)	KPK	1 (2)	No	Not present
OCA5	Mapped locus		Kausar[b] *et al* (117)	Punjab	1 (6)	No	-
MC1R	c.917G>A	p.(Arg306His)	Saleha *et al* (558)	NA	1 (3)	No	VUS/ benign (VCV000470716)
LYST	c.9827_9832 delATACAA	p.(Asn3276_Thr3277del)	Weisfeld-Adams *et al* (559)	NA	1 (3)	No	Pathogenic (VCV000180629)
HPS1	c.437G>A	p.(Trp146*)	Unpublished	Punjab (Shadra) [Bhatti Rajpoot]	1 (2)	No	Not present
HPS1	c.517C>T	p.(Arg173*)	Unpublished	KPK (Swabi) [Pashton, Yousafzai]	1 (1)	Chinese (402)	Pathogenic (VCV000937794)
HPS1	c.1342T>C	p.(Trp448Arg)	Yousaf *et al* (560)	Punjab	1 (5)	No	Likely pathogenic (VCV000690341)
HPS1	c.2009T>C	p.(Leu670Pro)	Unpublished	KPK (Peshawar) [Pashton]	1 (3)	No	Not present
HPS1	c.2056C>T	p.(Gln686*)	Yousaf *et al* (560)	Punjab	1 (4)	No	Likely pathogenic (VCV000690342)

HPS1	c.1397+1G>A	Affects splicing	Unpublished	Balochistan [Baloch]	1 (1)	No	Not present
HPS1	c.972delC	p.(Met325Trpfs*6)	Unpublished	Puunjab (RYK) [Bukhari Sayyed]	1 (3)	Japanese (401)	Pathogenic (VCV000005280)
HPS1	c.118-105_133 del121	p.(Leu40Thrfs*7)	Yousaf *et al* (560)	Punjab	1 (5)	No	-
HPS3	c.1509G>A	p.(Met503Ile)	Yousaf *et al* (560)	Punjab	1 (4)	No	Likely pathogenic (VCV000627005)
HPS4	c.276+5G>A	Affects splicing	Yousaf *et al* (560)	Sindh	1 (5)	No	Not present
HPS6	c.823C>T	p.(Pro275Ser)	Yousaf *et al* (560)	Punjab	1 (5)	No	Likely pathogenic (VCV000690344)
BLOC1S3	c.448delC	p.(Gln150Argfs*75)	Morgan *et al* (561)	NA	1 (6)	No	Pathogenic (VCV000001489)
BLOC1S6	c.232C>T	p.(Gln78*)	Yousaf *et al* (560)	Sindh	1 (2)	North Italian (562)	Pathogenic (VCV000030412)

Abbreviations: AJ&K, Azad Jammu and Kashmir; del, deletion; KPK, Khyber Pakhtunkhwa; NA, not available; RYK, Rahim Yar Khan; VUS, variant of uncertain significance. Study families are highlighted in red.

APPENDIX D: PRIMER PAIRS AND PCR CONDITIONS

Table D1 Primer pairs used for sequencing the *TYR* coding exons and associated intron-exon junctions

Primer	Primer Sequence 5' → 3'	Annealing temp (°C)	Amplicon size (bp)
TYR_Exon 1_F	TCAGCCAAGACATGTGATAATCA	60	992
TYR_Exon 1_R	TTATACCCTGCCTGAAGAAGTG		
TYR_Exon 2_F	CAACATTTCTGCCTTCTCCTA	55	888
TYR_Exon 2_R	CTGCCTAGAATATTTTAAACAGG		
TYR_Exon 3_F	GAATGAACAGGAGGGAACAC	58	470
TYR_Exon 3_R	TCTATTTAAATCCAATGAGCACGTT		
TYR_Exon 4_F	TTCTGGAGGTTCAAAACTCAATG	58	675
TYR_Exon 4_R	ACAAAATGGCCTATGTTAAGCAA		
TYR_Exon 5_F	TGTCTACTCCAAAGGACTGT	54	921
TYR_Exon 5_R	GGCACTTAGCTGGATGTGTT		

Table D2 Primer pairs used for sequencing the inherited eye disease gene variants identified in the study

Primer	Primer Sequence 5' → 3'	Annealing temp (°C)	Amplicon size (bp)
ALDH1A3_G414R_F	CAGTCTCCAATGGCAATGCA	60	359
ALDH1A3_R414R_R	CACACACAGTCAACTCACCA		
ALDH1A3_c172dupG_F	GGTGGACAAGATGGATAAGACG	58	300
ALDH1A3_c172dupG_R	GGCCAGTTCTGTCTTATAGCT		
ATOH7_c94delG_F	GGAAGCCGAAGAGTCTCTGG	57	592
ATOH7_c94delG_F	GCACTCCCCCACTGTAAACT		
BBS1_M390R_F	TTCCCCAGGCCTGTCTCTAT	60	539
BBS1_M390R_R	TCCGTCTTCCAGACGAGACT		
BBS5_ c196delA_F	ACACATATGACTTGCTGGGAC	56	676
BBS5_ c196delA_R	TGGGATTAATCAAAACAGGGGA		
CYP1B1_R390H_F	TCATCACTCTGCTGGTCAGG	58	388
CYP1B1_R390H_R	GAATTTTGCTCACTTGCTTTTC		
FRMD7_L148*_F	ATCTCAGCGTTTCATGGAGC	57	237

FRMD7_L148*_R	TGCAGCAGAACTTGGAGACT		
FYCO1_Q736*_F	CTTAGCTGGCTCTGCACCTT	58	363
FYCO1_Q736*_R	CCAGATGGCAGAGAAGAAGG		
HPS1_W146*_F	TGCTTGTGCCTTCATTCATT	56	388
HPS1_W146*_R	CCCCACTCCACAGTTAGAGC		
HPS1_R173*_F	AGAGTAGAATGCCAGCAGCTT	64	243
HPS1_R173*_R	AGTGAATGTCCCCACTAGCC		
HPS1_L670P_F	ACAGGATGCAAAGGCAGACT	60	240
HPS1_L670P_R	CCACAGCCTCACTCCTGAAT		
HPS1_c972delC_F	ACAATGGAGCTGAGGGACAG	58	468
HPS1_c972delC_R	TTAGGATGAAGGGGTGTTGC		
HPS1_c1391+1_F	GGGCATTACAGCAGAAGGGA	61	435
HPS1_c1391+1_R	CAAAAGTGAGCCCGGATCCT		
INPP5E_ Q627*_F	AGGGTTGGCTTCCTTCCTG	58	392
INPP5E_ Q627*_R	GCCCTGGGTGTCCTCTTAAA		
LRP5_T359R_F	GCCTGGCTGAGTATTTCCCT	58	580
LRP5_T359R_R	CCAGAATGACAGGTCCAGGT		
MKKS_H84Y_F	TCGACAACCACAGGTCTCAG	60	240
MKKS_H84Y_R	TCCTCAGCTCTGCTCAGTCA		
OCA2_ V443I_F	GGACTGGAATGCAGTGAGCT	64	321
OCA2_ V443I_R	TCTACGAGCCTGCTCACTCT		
OCA2_ D486Y_F	CTTCCTCAGCTCTTGGTTGG	60	230
OCA2_ D486Y_R	ATACGAGCAAGCGCCTTAAA		
OCA2_ R588W_F	TGGTTTTTGTCGTGCAACAA	64	578
OCA2_ R588W_R	CATCGACTGTGTGGGGAACA		
OCA2_ L674V_F	ACAAATACGCAGTGCTGTCAG	62	320
OCA2_ L674V_R	TTCTGGGTGCCATCTGGTTG		
OCA2_ S736L_F	TACACCTGTGAGTGCAGCAG	56	567
OCA2_ S736L_R	GTAAATGAGTGCTGCAGGCG		
OCA2_ Ala787V_F	CCTGAAAAATTCCATGAAGGAG	58	196
OCA2_ Ala787V_R	CAAATCAAAGCCTGTGAGATGA		
OCA2_ S820P_F	TCTGGAGGGGAATCTTGAGT	56	551
OCA2_ S820P_R	TGCACACAATGGAGGATGTC		
OCA2_c1045-15_F	AGTTCTGTGCACGATCTGGA	58	184
OCA2_c1045-15_R	TATGTGTCTGTGGGGTGTCC		
OCA2_c408delAA_F	GTCTCCCTAGGCCCAGGTAA	62	501
OCA2_c408delAA_R	AGAGTGCACCACTGTCTGTG		
PAX6_R240*_F	TGAATCACAAAGTGTGAAACTGC	64	206
PAX6_R240*_R	AGGTGGGAACCAGTTTGATG		

SCAPER_c2236dup_F	TTTGATCTCATGCTCCACTGT	56	401
SCAPER_c2236dup_R	AGCCCTTTAGACTATATGACCCT		
SDHD_E69K_F	CTGCCTGTCAGTTTGGGTTA	58	243
SDHD_E69K_R	TTGCCAGTGACCATGAAGAG		
TDRD7_c2469delG_F	GGATTTTAGCAAGGGTTTTGG	60	396
TDRD7_c2469delG_R	GAGAAAGCGCTCGTATGGTC		

APPENDIX E: PUBLICATIONS RELATING TO THIS RESEARCH

First/joint first author

Lin S, Haralka GV, Hameed A, Reham HM, Yasin M, Muhammad N, Khan S, Baple EL, Crosby AH, Saleha S. Novel mutations in ALDH1A3 associated with autosomal recessive anophthalmia/microphthalmia, and a review of literature. *BMC Medical Genetics* 2018 Sep 10; 19 (1):160

Lin S, Sanchez-Bretaño A, Leslie JS, Williams KB, Lee H, Thomas NS, Callaway J, Deline J, Ratnayaka JA, Barelle D, Schmitt MA, Norman CS, Hammond S, Harlalka GV, Ennis S, Cross HE, Crosby AH, Baple EL, Self JE. Evidence that the Ser192Tyr/Arg402Gln in cis Tyrosinase gene haplotype is a disease-causing allele in oculocutaneous albinism type 1B (OCA1B). *In submission to npj Genomic Medicine*

Lin S*, Fasham J*, Al-Hijawi F*, Qutob N, Gunning A, Leslie JL, McGavin L, Ubeyratna N, Baker W, Zeid R, Turnpenny PD, Crosby AH, Baple EL, Khalaf-Nazzal R. Consolidating biallelic SDHD variants as a cause of mitochondrial complex II deficiency. Eur J Hum Genet 2021 May 20. doi: 10.1038/s41431-021-00887-w. Online ahead of print

Fasham J*, Arno G*, **Lin S***, Xu M, Carss KJ, Hull S, Lane A, Robson AG, NIHR Bioresource Rare Diseases Consortium, Wenger O, Self JE, Harlalka GV, Salter CG, Schema L, Moss TJ, Cheetham ME, Moore AT, Raymond FL, Chen R, Baple EL, Webster AR, Crosby AH; NIHR Bioresource Rare Diseases Consortium. Delineating the expanding phenotype associated with *SCAPER* gene mutation. Am J Med Genet A 2019 Aug; 179(8):1655-1671

Co-author

Shakil M, Harlalka GV, Ali S, **Lin S**, D'Atri I, Hussain S, Nasir A, Shahzad M, Ullah M, Self JE, Baple E, Crosby A, Mahmood S. Tyrosinase (TYR) gene sequencing and literature review reveals recurrent mutations and multiple founder gene mutations as causative of oculocutaneous albinism (OCA) in Pakistani families. Eye (Lond) 2019 Aug; 33(8): 1339-1346

Khan S, **Lin S**, Harlalka GV, Ullah A, Shah K, Khalid S, Mehmood S, Hassan MJ, Ahmad W, Self JE, Crosby AH, Baple EL, Gul A. BBS5 and INPP5E mutations associated with ciliopathy disorders in families from Pakistan. Ann Hum Genet 2019 Nov; 83(6): 477-482

Arshad MW, Harlalka GV, **Lin S**, D'Atri I, Mehmood S, Shakil M, Hassan MJ, Chioza BA, Self JE, Ennis S, O'Gorman L, Norman C, Aman T, Ali SS, Kaul H, Baple EL, Crosby AH, Ullah MI, Shabbir MI. Mutations in TYR and OCA2 associated with oculocutaneous albinism in Pakistani families. *Meta Gene* 2018 Sept; 17: 48-55